

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 1/21, 5/10, 15/11, 15/12, 15/63, A61K 38/16, 38/17, C07K 14/00, 14/435, 16/00, G01N 33/50	A1	(11) International Publication Number: WO 00/06698 (43) International Publication Date: 10 February 2000 (10.02.00)
(21) International Application Number: PCT/US99/17130 (22) International Filing Date: 29 July 1999 (29.07.99) (30) Priority Data: 60/094,657 30 July 1998 (30.07.98) US 60/095,486 5 August 1998 (05.08.98) US 60/095,455 6 August 1998 (06.08.98) US 60/095,454 6 August 1998 (06.08.98) US 60/096,319 12 August 1998 (12.08.98) US (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): KOMATSOULIS, George, A. [US/US]; 9518 Garwood Street, Silver Spring, MD 20901 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Ol- ney, MD 20832 (US). DUAN, Roxanne, [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). MOORE, Paul, A. [GB/US]; 19005 Leatherbark Drive, Germantown, MD 20874 (US). SHI, Yanggu [CN/US]; Apartment 102,		437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David [US/US]; 3142 Quesada Street, N.W., Washington, DC 20015 (US). WEI, Ying-Fei [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; Apartment 115, 410 Van Dyke Street, St. Paul, MN 55119-4321 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). MUCENSKI, Michael [US/US]; 3265 Mandale Drive, Cincinnati, OH 45239 (US). (74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>With an indication in relation to deposited biological</i> <i>material furnished under Rule 13bis separately from the</i> <i>description.</i>
(54) Title: 98 HUMAN SECRETED PROTEINS (57) Abstract <p>The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

98 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and
5 their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or
10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic
15 reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

20 Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in
25 secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins
30 include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of

the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

5

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

15

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

25

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

30

In specific embodiments, the polynucleotides of the invention are less than 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, or 7.5 kb in length. In a further embodiment, polynucleotides of the invention comprise at least 15 contiguous nucleotides of the coding sequence, but do not comprise all or a portion of any intron.

5 In another embodiment, the nucleic acid comprising the coding sequence does not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene in the genome).

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone
10 deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a
15 molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone containing all or most of the sequence for SEQ ID
20 NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of
25 microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an
30 overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's

solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions.

- 5 Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M
- 10 NaH_2PO_4 ; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

- Note that variations in the above conditions may be accomplished through the
- 15 inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above,
- 20 due to problems with compatibility.

- Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid
- 25 molecule containing a poly (A) stretch or the complement thereof (e.g., practically, any double-stranded cDNA clone).

- The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of
- 30 single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA

that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA
5 backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids
10 joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more
15 detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of
20 modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a
25 heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation,
30 iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

(See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene is a human glycoprotein-associated amino acid transporter (See, e.g., Genbank Accession No. embICA10198.11 (AJ130718); all references available through this accession are hereby incorporated by reference herein). Amino acid transport across cellular membranes is mediated by multiple transporters with overlapping specificities. The transport system L, which mediates Na⁺-independent exchange of large neutral amino acids, consists of a novel amino acid permease-related protein (LAT1 or AmAT-L-1c) which for surface expression and function requires formation of disulfide-linked heterodimers with the glycosylated heavy chain of the h4F2/CD98 surface antigen. h4F2hc also associates with other mammalian light chains, e.g. y+LAT1 from mouse and human which are

approximately 48% identical with LAT1 and thus belong to the same family of glycoprotein-associated amino acid transporters.

The novel heterodimers form exchangers which mediate the cellular efflux of cationic amino acids and the Na⁺-dependent uptake of large neutral amino acids.

- 5 These transport characteristics and kinetic and pharmacological fingerprints identify them as y⁺L-type transport systems. mRNA encoding y⁺LAT1 is detectable in most adult tissues and expressed at high levels in kidney cortex and intestine. This indicates that the y⁺LAT1-4F2hc heterodimer, besides participating in amino acid uptake/secretion in many cell types, is the basolateral amino acid exchanger involved
10 in transepithelial reabsorption of cationic amino acids; hence, its defect might be the cause of the human genetic disease lysinuric protein intolerance.

The gene encoding the disclosed cDNA is believed to reside on chromosome 14. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 14.

- 15 Preferred polypeptides comprise the following amino acid sequence:
LALYSALFSYSGWDTLN (SEQ ID NO: 237), VTEEIKNPERNLPL (SEQ ID NO: 238), IGISMPIVT (SEQ ID NO: 239), IYILTNVAYYTVL (SEQ ID NO: 240), SDA
VAVTFADQ (SEQ ID NO: 241), VALSCFGGLNASI (SEQ ID NO: 242), SRLFFV
GSREGHLPD (SEQ ID NO: 243), SFSYWFFVGLS (SEQ ID NO: 244), VGQLYLR
20 WKEP (SEQ ID NO: 245), RPRPLKLSVFFPIVFC (SEQ ID NO: 246), DTINSLIGI
(SEQ ID NO: 247), LLAAACICLLTFINCA YVKWGTLVQDIFTYAKVLALIAVI
VAGIVRLGQGASTHFENSFEGSSFAVGDIALLALYSALFSYSGWDTLNYVTEEI
KNPERNLPLSIGISMPIVTIYILTNVAYYTVLDMRDILASDAVAVTFADQIFGIF
NWIPLSVALSCFGGLNASIVAASRLFFVGSREGHLPDAICMIHVERFTPVPSLL
25 FNGIMALIYLCVEDIFQLINY SFSYWFFVGLSIVGQLYLRWKEPDRPRPLKLS
VFFPIVFCCTIFLVAVPLYSDTINSLIGIAIALSGLPFYFLIIRVPEHKRPLYLRRI
VGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN (SEQ ID NO: 249)
and/or ATALPPKIVGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN
(SEQ ID NO: 248). Polynucleotides encoding these polypeptides are also provided.

- 30 Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 monocyte cells to calcium. Thus, it is likely that the product of this gene is involved in a signal

transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both THP-1 monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocytes.

5 This gene is expressed primarily in endothelial cells and brain, and, to a lesser extent, in a wide variety of human tissues.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
10 not limited to, disorders of the neural or gastrointestinal systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulation system or central nervous system, expression of this gene at significantly higher or
15 lower levels is routinely detected in certain tissues or cell types (e.g., neural, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not
20 having the disorder.

 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 124 as residues: Glu-102 to Asn-110, Arg-256 to Leu-266, Pro-316 to Trp-328, Pro-331 to Arg-336, Met-350 to Gly-358.

Polynucleotides encoding said polypeptides are also provided.

25 The tissue distribution in brain combined with its homology to a amino acid transporter and biological activity of increasing ion flux in monocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the
30 "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,

Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1550 of SEQ ID NO:11, b is an integer of 15 to 1564, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The gene encoding the disclosed cDNA is believed to reside on the X chromosome. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for the X chromosome.

Preferred polypeptides of the invention comprise the following amino acid sequence:

AARGSGVRDPLEEAVCPFSDLQLHAGRTTALFKAVRQGHLSLQRLLLSFVCL
CPAPRGGAYRGRQASLSCGGLHPVRASRLCLPKQAWAMAGAPPPVSLPPCS
LISDCCASNQRDSVG (SEQ ID NO: 250). Polynucleotides encoding these
polypeptides are also provided.

5 This gene is expressed primarily in cord blood cells, and, to a lesser extent, in
frontal lobe of the brain.

 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
10 not limited to, developmental, reproductive, hematopoietic or neural disorders.
Similarly, polypeptides and antibodies directed to these polypeptides are useful in
providing immunological probes for differential identification of the tissue(s) or cell
type(s). For a number of disorders of the above tissues or cells, particularly of the
immune and central nervous systems, expression of this gene at significantly higher or
15 lower levels is routinely detected in certain tissues or cell types (e.g., immune,
developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,
amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue
or cell sample taken from an individual having such a disorder, relative to the
standard gene expression level, i.e., the expression level in healthy tissue or bodily
20 fluid from an individual not having the disorder.

 Preferred polypeptides of the present invention comprise immunogenic
epitopes shown in SEQ ID NO: 125 as residues: His-56 to Gln-65, Leu-80 to Ile-85.
Polynucleotides encoding said polypeptides are also provided.

 The tissue distribution in cord blood cells indicates polynucleotides and
25 polypeptides corresponding to this gene are useful for the treatment and diagnosis of
hematopoietic related disorders such as anemia, pancytopenia, leukopenia,
thrombocytopenia or leukemia since stromal cells are important in the production of
cells of hematopoietic lineages. Representative uses are described in the "Immune
Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19,
30 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo
culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or
chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1743 of SEQ ID NO:12, b is an integer of 15 to 1757, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in human T cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T cell lymphoma, immunodeficiencies, in addition to other immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
5 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample
10 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 126 as residues: Met-1 to Phe-10. Polynucleotides
15 encoding said polypeptides are also provided.

The tissue distribution in human T cell lymphomas indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14,
20 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting
25 immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory
30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1359 of SEQ ID NO:13, b is an integer of 15 to 1373, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

The protein product of this clone shares sequence homology with the C-terminus of a human N-acetylglucosamine-phosphate mutase (See, e.g., Genbank Accession No. gb|AAC72409.11 (AF102265); all references available through this accession are hereby incorporated by reference herein.) Hofmann, et al. (Eur. J. Biochem. 221:741-747 (1994)) studied the N-acetylglucosamine-phosphate mutase of *Saccharomyces cerevisiae* and showed it to be essential for viability. A *S. cerevisiae* agm1 deletion mutant progressed through only approximately five cell cycles to form a 'string' of undivided cells with an abnormal cell morphology resembling

glucosamine auxotrophic mutants. Expression of the AGM1 gene on a multi-copy plasmid led to a significantly increased N-acetylglucosamine-phosphate mutase activity. Unlike over-expression of the *S. cerevisiae* AGM1 gene in a phosphoglucumutase (pgm1 delta/pgm2 delta) double deletion mutant which could
5 restore phosphoglucumutase activity, over-expression of the PGM2 gene encoding the major isoenzyme of phosphoglucumutase did not increase N-acetylglucosamine-phosphate-mutase activity and did not restore growth of agm1 deletion mutant cells. These observations indicate that the different hexosephosphate mutases of *S. cerevisiae* have partially overlapping substrate specificities but, nevertheless, distinct
10 physiological functions. The human N-acetylglucosamine-phosphate-mutase is expected to share at least some biological activities with the Agm1 protein.

Preferred polypeptide fragments of the invention comprise the following amino acid sequences: LSKAFLDSPNRLLAVEMNTDHLRLTVPNGIGALKLRXM
EHYFSQGLSVQLFNDGSKGKLNHL CGAD FVKSHQKPPQGMEIKSNERCCSFD
15 GDADRIVYYYHDADGHFHLIDGDKIATLISSFLKELLVEIGESLNIGVVQTAYA
NGSSTRYLEEVMKVPVYCTKTGVKHLHHKAEFDIGVYFEANGHG TALFST
AVEMKIKQSAEQLEDKKRKA AKMLENIIDLFNQAAGDAISDMLVIEAILALK
GLTVQQWDALYTDLPNRQLKVQVADRRVISTTXAERQAVTPPGLQEAINDL
VKKYKLSRAFVRPSGTEDVVRVYAEADSQESADHLAHEVSLAVFQLAGGIGE
20 RPQPGF (SEQ ID NO: 251), LSKAFLDSPNRLLAVEMNTDHLRLTV (SEQ ID
NO: 252), PNGIGALKLRXMEHYFSQGLSVQLFNDG (SEQ ID NO: 253), SKGKL
NHL CGAD FVKSHQKPPQGMEIKS (SEQ ID NO: 254), NERCCSFDGDADRIV
YYYHDADGHFHLI (SEQ ID NO: 255), DGDKIATLISSFLKELLVEIGESLNIGV
(SEQ ID NO: 256), VQTAYANGSSTRYLEEVMKVPVYCTKTG (SEQ ID NO:
25 257), VKHLHHKAEFDIGVYFEANGHG TALFS (SEQ ID NO: 258), TAVEMK
IKQSAEQLEDKKRKA AKMLENI (SEQ ID NO: 259), IDLFNQAAGDAISDM
LVIEAILALKGLT (SEQ ID NO: 260), VQQWDALYTDLPNRQLKVQVADRR
VIST (SEQ ID NO: 261), TXAERQAVTPPGLQEAINDLVKKYKLSR (SEQ ID
NO: 262), AFVRPSGTEDVVRVYAEADSQESA (SEQ ID NO: 263), and/or DH
30 LAHEVSLAVFQLAGGIGERPQPGF (SEQ ID NO: 264). Polynucleotides encoding
these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

5 This gene is expressed primarily in fetal brain, and, to a lesser extent, in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental disorders, particularly of the central nervous system.

10 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central and peripheral nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or

15 cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 127 as residues: Asn-36 to Lys-42, Lys-53 to Gln-60, Ile-64 to Ala-77, Ala-128 to Tyr-135, Lys-184 to Ala-199, Leu-245 to Leu-250. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of N-acetylglucosamine-phosphate mutase in fetal

25 brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to

30 the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital

malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3726 of SEQ ID NO:14, b is an integer of 15 to 3740, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in human stomach and stomach tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the gastrointestinal system, particularly cancer or ulcers of stomach tissue. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the

tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of the stomach indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment and intervention of these tumors, in addition to other tumors where expression has been indicated. Additionally, the protein product of this gene may play a role in the normal function of the stomach and/or digestive system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1182 of SEQ ID NO:15, b is an integer of 15 to 1196, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 6

Preferred polypeptides of the invention comprise the following amino acid sequences:

FEIALPRESNITVLIKLGTPTLAKPCYIVISKRHITMLSIKSGERIVFTFSCQSPE
NHFVIEIQKNIDCMSGPCPFGEVQLQPSTSLPTLNRTFIWDVKAHKSIGLELQ
FSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTFCSNGTVSRIKM (SEQ ID
NO: 266), and/or GTRAAPGLGAWGRRSPPSFSPRRPGVMAGLNCGVSIAL
5 LGVLLLGAARLPRGAFAFEIALPRESNITVLIKLGTPTLAKPCYIVISKRHITM
LSIKSGERIVFTFSCQSPENHFVIEIQKNIDCMSGPCPFGEVQLQPSTSLPTLN
RTFIWDVKAHKSIGLELQFSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTF
CSNGTVSRIKMQEGVKMALHLPWFHPRNVSGFSIANRSSIKRLCIESVFEGEGS
ATLMSANYPEGFPEDELMTWQFVVAHLRASVSFLNFNLSNCERKEERVEYY
10 IPGSTTNPEVFKLEDKQPGNMAGNFNLSLQGCDQDAQSPGILRLQFQVLVQH
PQNESNKIYVVDLSNERAMSLTIEPRPVKQSRKFVPGCFVCLESRTCSSNLTLT
SGSKHKISFLCDDLTRLWMNVEKP (SEQ ID NO: 265). Polynucleotides encoding
these polypeptides are also provided.

This gene is expressed primarily in placenta, and to a lesser extent in, prostate
15 and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, male and female infertility, and associated disorders of the
20 reproductive system. Similarly, polypeptides and antibodies directed to these
polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the reproductive system, expression of this gene at
significantly higher or lower levels is routinely detected in certain tissues or cell types
25 (e.g., reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,
amniotic fluid, seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative
to the standard gene expression level, i.e., the expression level in healthy tissue or
bodily fluid from an individual not having the disorder.

30 The tissue distribution of this gene in the prostate, placenta and ovary
indicates that polynucleotides and polypeptides corresponding to this gene are useful
for treatment, prevention, and/or diagnosis of male or female infertility, endocrine

disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia and prostate cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within placental tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2195 of SEQ ID NO:16, b is an integer of 15 to 2209, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares homology with the human and rat HNK-1 sulfotransferase protein (See, e.g., Genbank Accession Nos. gblAAB88123.11 (AF022729) and gil2921306lgblAAC04707.11 (AF033827); all references available through these accessions are hereby incorporated herein by reference.) Ong E, et al. (J Biol Chem. 273(9):5190-5 (1998)) have characterized the human HNK-1 sulfotransferase, and show that it is involved in the synthesis of the HNK-1

carbohydrate epitope which is expressed on various adhesion molecules in the nervous system and on immune cells (e.g., natural killer cells) and is suggested to play a role in cell-cell and cell-substratum interactions. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with HNK-1 sulfotransferase proteins. Such activities are known in the art, some of which are described elsewhere herein, or in, for example, Bakker, et al., J Biol Chem. 272:29942-6 (1997), incorporated herein by reference. Based on sequence similarity between sulfotransferases, a consensus sequence for the active site was developed (Ong, et al., supra). The consensus pattern is as follows:

5 xxRPDzzzz, where x represents hydrophobic amino acid residues and z represents any amino acid residue. Therefore,

Preferred polypeptides of the invention comprise the following amino acid sequences: FVRDPFVRL (SEQ ID NO: 267), FLFVRDPFVRLIS (SEQ ID NO: 268), FLFVRDPFVRLISAF (SEQ ID NO: 269), and/or YLHTSFSRPHTGPPLPTPG

15 PDRDRELTADSDVDEFLDKFLSAGVKQSDLPRKETEQPPAPGSMEENVRGY
DWSPRDARRSPDQGRQQAERRSVLRGFCANSSLAFPTKERAFDDIPNSELSHL
IVDDRHGAIYCYVPKVACTNWKRVMIVLSGSLLHRGAPYRDPLRIPREHVH
NASAHLTFNKFWRRYGKLSRHLMKVKLKKYTKFLFVRDPFVRLISAFRSK
FELENEEFYRKFAVPMLRLYANHTSLPASAREAFRAGLKVSFANFIQYLLDPH

20 TEKLAPFNEHWRQVYRLCHPCQIDYDFVGKLETLEDAQAQLLQLLQVDRQ
LRFPPSYRNRTASSWEEDWFAKIPLAWRQQLYKLYEADFVLFGYPKPENLL
RD (SEQ ID NO: 270). Polynucleotides encoding these polypeptides are also provided.

Further preferred are the sulfotransferase active site polypeptides listed above,

25 and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the sulfotransferase active site polypeptides. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the sulfotransferase active site polypeptides, wherein the

30 total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to sulfotransferase proteins. The nucleotides sequence of this gene was found to be

homologous to the human hypoxanthine guanine phosphoribosyl transferase 2 cDNA which is known to be involved in the purine salvage pathway resulting in the maintenance of homeostatic levels of uric acid (See Genbank Accession No.T30127).

The gene encoding the disclosed cDNA is believed to reside on chromosome

- 5 7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

This gene is expressed to a very high level in HL-60 myelogenous leukemia cell lines, and to a lesser extent, in most cell types of the immune system.

- Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, myelopoiesis, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, metabolic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an
20 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 130 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62,
25 Lys-79 to Ala-93, Gly-95 to Asp-105, Ser-107 to Val-127, Gly-193 to Leu-200, Lys-218 to Ser-227, Lys-234 to Thr-239, Pro-366 to Asp-379, Pro-406 to Asp-414. Polynucleotides encoding said polypeptides are also provided.

- The tissue distribution in HL-60 myelogenous leukemia cell lines and homology to HNK-1 sulfotransferase proteins indicates that polynucleotides and
30 polypeptides corresponding to this gene are useful for the diagnosis, prevention and/or treatment of a variety of immune system disorders, including but not limited to, those involving the HNK-1 carbohydrate epitope, (e.g. HNK-1 as an auto-antigen

in peripheral demyelinating neuropathy). Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, expression of this gene product in tonsils indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the homology to a conserved purine metabolism protein may suggest that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sach's Disease, phenylketonuria, galactosemia, porphyrias, Hurler's syndrome, and various urogenital disorders related to metabolic conditions, particularly Lesch-Nyhan syndrome. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1760 of SEQ ID NO:17. b is an integer of 15 to 1774, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:17, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS) promoter element. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway, a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: KLVRLQVPVRNSRVDPRVRSKIGSRRWMLQLI
MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVPYLYFDMGAAVLCA
SFMSFGVKRRWFALGAALQLAISTYAAAYIGGYVHYGDWLKVRMYSRTVAII
GGFLVLASGAGELYRRKPRSRLQSTGQVFLGIYLCVAYSLQHSKEDRLA
YLNHLPGGELMIQLFFVLYGILALAFLSGYVTLAAQILAVLLPPVMLLIDG
NVAYWHNTRRVEFWNQMKLLGESVGIFGTAVILATDG (SEQ ID NO: 271).

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVPYLYFDMGAAVLCASFMSFGVKRRWFALGAALQLAISTYAAAYIGGYVHYGD

WLKVRMYSRTVAIIGGFLVLASGAGELYRRKPRSRSLQSTGQVFLGIYLICVA
YSLQHSKEDRLAYLNHLPGGELMIQLFFVLYGILAPGLSVRLLRDPRCPDPGC
TAAPCHAAH (SEQ ID NO: 272). Polynucleotides encoding these polypeptides are also provided.

5 The gene encoding the disclosed cDNA is believed to reside on chromosome 17. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 17.

 This gene is expressed primarily in endometrial tumors, and to a lesser extent, in T-cells, pituitary and to a certain extent in most cell types.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, female reproductive, immune, or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
15 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and/or immune systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, lymph, serum,
20 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 Preferred polypeptides of the present invention comprise immunogenic
25 epitopes shown in SEQ ID NO: 131 as residues: Ala-27 to Asp-34, Tyr-116 to Leu-125. Polynucleotides encoding said polypeptides are also provided.

 The tissue distribution predominantly in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of a range of immune and/or reproductive
30 disorders including endometriosis, endometritis, and endometrioma. Similarly, the tissue distribution in T-cells and the ability of supernatants expressing this gene to stimulate the GAS promoter element in T-cells indicates polynucleotides and

polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

10 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in pituitary indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of the Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-

hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1660 of SEQ ID NO:18, b is an integer of 15 to 1674, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of the myeloid leukemia cell line AML-193 to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of myeloid leukemia cells, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating myeloid leukemia cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: SNEILLSFPQNYIQLNGSLIHGLWNLASLFS NLCLFVLMPFAFFLESEGFAGLKKGIRARILETLVMLLLLALLILGIVWVAS ALIDNDAASMESLYDLWEFYLPYLYSCISLMGCLLLLCTPVGLSRMFTVMG HLLVKPTILEDLDEQIYIITLEEEALQRRNLGLSSSV EYNIMELEQELENVKTL KTKLERRKKASAWERNLVYPVAVMVLIIETSSISVLLVACNILCLLVDETAM PKGTRGPGIGNASLSTFGFVGAALEIILIFYLMVSSVVG FYSLRFFGNFTPKKD DTTMTKIIGNCVSILVLSSALPVMSRTLGITRFDLLGDFGRFNWLGNFYIVLS YNLLFAIVTTLCLVRKFTSAVREELFKALGLHKLHLPNTSRDSETAKPSVNGH

QKAL (SEQ ID NO: 273). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in fetal heart, and to a lesser extent, in colon and the adult pulmonary system.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, heart, lung and digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological
10 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, pulmonary and digestive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid,
15 serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
20 epitopes shown in SEQ ID NO: 132 as residues: Glu-67 to Asn-74, Glu-88 to Asn-93, Lys-95 to Ala-107, Ala-147 to Arg-153, Phe-197 to Thr-205, Pro-292 to His-308. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal heart, colon and pulmonary tissues and the biological activity in increasing the permeability of the plasma membrane of the
25 myeloid leukemia cell line AML-193 to calcium, likely indicating that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of myeloid leukemia cells, indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention, and/or detection of a range of disorders including
30 a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, myocardial infarction, myocarditis, ischemia, thrombosis, coronary artery disease,

arteriosclerosis, and/or atherosclerosis; pulmonary edema and embolism, bronchitis and/or cystic fibrosis; Crohn's Disease and/or colon cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2004 of SEQ ID NO:19, b is an integer of 15 to 2018, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

The protein product of this clone shares sequence homology with the human MaxiK channel beta 2 subunit (See, Genbank Accession No.

gb|AAD23380.1|AF099137_1 (AF099137); all references available through this accession are hereby incorporated herein by reference), which is believed to be a modulatory subunit of the voltage and Ca^{2+} activated K^{+} (MaxiK) channel.

Additionally, this protein shares homology to the human calcium-activated potassium

channel beta subunit, which, when combined with its corresponding alpha subunit and modulating peptide, are believed to be useful in treating asthma, angina, hypertension, incontinence, migraine, irritable bowel syndrome (IBS). The subsequent heteromultimer that forms upon combining the alpha, beta, and modulator subunits are also thought to be useful in preventing premature labour, preventing and treating cerebral ischemia, inducing pain modulation and decreasing neurogenic inflammation in a patient (See GeneSeq Accession No. R85306).

Preferred polypeptides comprise the soluble domain which consists of the following amino acid sequence: RSYMQSVWTEESQCTLLNASITETFNCSFSCGP
10 DCWKLSQYPCLQVYVNLTSSEKLLLYHTEETIKINQKCSYIPKCGKNFEESM
SLVNVVMENFRKYQHFSCYSDPEGNQKSVILTKLYSSNVLFHSLFWPTCMMA
GGVAIVAMVKLTQYLSLLCERIQRINR (SEQ ID NO: 274). Polynucleotides encoding these polypeptides are also provided. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with modulatory subunits of voltage and Ca²⁺ activated K⁺ channel proteins. Such activities are known in the art, some of which are described in Wallner, et al., PNAS 96:4137-4142 (1999), incorporated herein by reference.

This gene is expressed primarily in adrenal gland tumor, and to a lesser extent, in Hodgkin's lymphoma.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endocrine and immune disorders, particularly Hodgkin's Lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and/or endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,
30 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression

level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 133 as residues: Trp-25 to Gln-30, Pro-50 to Gln-57,

5 Pro-93 to Glu-101, Arg-114 to Cys-121, Ser-123 to Gln-129, Ile-177 to Arg-182.

Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in adrenal gland tumor and its identification as the modulatory subunit of the voltage and Ca^{2+} activated K^{+} (MaxiK) channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the

10 detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Alternatively, expression in

15 proliferative immune tissues combined with its homology to a novel human K channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein.

20 Briefly, the expression of this gene product in Hodgkin's lymphoma indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune

25 responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory

30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity, immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2084 of SEQ ID NO:20, b is an integer of 15 to 2098, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

The translation product of this gene shares homology with collagen and collagen like proteins (See, e.g., Genbank Accession Nos. gil2920535|gb|AAC39658.11 (AF018081) and gil2384942|gb|AAB69961.11 (AF022985); all references available through these accession numbers are hereby incorporated by reference herein). Additionally, it has been determined that this gene has homology to the human Kruppel related zinc finger protein (HTF10) which is known to be important as a transcription factor, particularly in development (See Genebank Accession No.L11672).

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AFAHLQLGPMWKLWRAEEGAAALGGALFLLL
5 FALGVRQLLKQRRPMGFPPGPPGLPFIGNIYSLAASSELPHVYMRKQSQVYG
EVQPRRAPGREGRQAGPGWPGPSWLDLWPPLGRLVGTSPCAGCPLRDTRFPG
LEGRS PRRRAPLQGEPRPCR (SEQ ID NO: 275). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in human erythroleukemia cell line (HEL),
10 serum induced smooth muscle, and to a lesser extent in human 8 week whole embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia, musculoskeletal, or developmental disorders. Similarly,
15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system and muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune,
20 musculoskeletal, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 134 as residues: Leu-30 to Gly-38, Arg-67 to Val-72, Val-76 to Ala-89, Pro-118 to Arg-123, Gly-129 to Ala-136, Leu-138 to Arg-146. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human erythroleukemia cell line (HEL), and serum
30 induced smooth muscle, and the shared homology with collagen and collagen like proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for disorders of hematopoietic or muscular systems, such as leukemia and

muscular dystrophy. Additionally, the shared homology with collagen proteins would suggest that this protein may also be important in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and
5 chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.).

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents
10 that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines:
immunostimulating/immunosuppressant activities (e.g., for treating human
15 immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating
20 infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy
25 procedures. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
5 formula of a-b, where a is any integer between 1 to 1732 of SEQ ID NO:21, b is an integer of 15 to 1746, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

10 A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRVRIGLTLLLCVLLSLASASSDEEGSQD
ESLGFQDYFDIR (SEQ ID NO: 276). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome
15 8. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 8.

This gene is expressed primarily in dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at
25 significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 135 as residues: Ser-22 to Ser-41, Glu-43 to Thr-50,

Ser-63 to Leu-68, Ser-71 to Gly-84, Ser-96 to Gly-114. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation

modulating activity or induction of other cytokines:
immunostimulating/immunosuppressant activities (e.g., for treating human
immunodeficiency virus infection, cancer, autoimmune diseases and allergy);
regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to
5 chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or
nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for
control of fertility); chemotactic and chemokinetic activities (e.g., for treating
infections, tumours); haemostatic or thrombolytic activity (e.g., for treating
haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating
10 septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other
hyperproliferative disease; for regulation of metabolism, behaviour, and many others.
Also contemplated is the use of the corresponding nucleic acid in gene therapy
procedures. Furthermore, the protein may also be used to determine biological
activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to
15 identify agents that modulate their interactions, in addition to its use as a nutritional
supplement. Protein, as well as, antibodies directed against the protein may show
utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
available and accessible through sequence databases. Some of these sequences are
20 related to SEQ ID NO:22 and may have been publicly available prior to conception of
the present invention. Preferably, such related polynucleotides are specifically
excluded from the scope of the present invention. To list every related sequence is
cumbersome. Accordingly, preferably excluded from the present invention are one or
more polynucleotides comprising a nucleotide sequence described by the general
25 formula of a-b, where a is any integer between 1 to 2862 of SEQ ID NO:22, b is an
integer of 15 to 2876, where both a and b correspond to the positions of nucleotide
residues shown in SEQ ID NO:22, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

30 A preferred polypeptide variant of the invention comprises the following
amino acid sequence: MARGSLRRLRLVLGLWLALLRSVAGEQAPGTAPC
SRGSSWSADLDKCMDCTSCPLPA ALAHPWGRSEPDRLAGAAFWLFGLE

TMPQE REVHHPHRGDRRRGLPSCGADPVTMCPLPAGARPLIIHSSILEPVSAS
QTRREPSSSNHK GGGGR (SEQ ID NO: 277). Polynucleotides encoding these
polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome
5 16. Accordingly, polynucleotides related to this invention are useful as a marker in
linkage analysis for chromosome 16.

This gene is expressed primarily in tumor growth factor or lipopolysaccharide
treated bone marrow stroma, epithelioid sarcoma, umbilical vein endothelial cells, and
to a lesser extent, in other tissues.

10 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, hematopoiesis or immune disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological
15 probes for differential identification of the tissue(s) or cell type(s). For a number of
disorders of the above tissues or cells, particularly of the hematopoietic,
integumentary, or immune systems expression of this gene at significantly higher or
lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic,
immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,
20 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken
from an individual having such a disorder, relative to the standard gene expression
level, i.e., the expression level in healthy tissue or bodily fluid from an individual not
having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
25 epitopes shown in SEQ ID NO: 136 as residues: Pro-35 to Trp-42, Pro-65 to Asp-72,
Thr-86 to Phe-93, Ile-97 to Glu-103. Polynucleotides encoding said polypeptides are
also provided.

The tissue distribution in tumor growth factor or lipopolysaccharide treated
bone marrow stroma, epithelioid sarcoma, and umbilical vein endothelial cells
30 indicates that polynucleotides and polypeptides corresponding to this gene are useful
for the diagnosis, prevention, and/or treatment of a variety of immune system
disorders. Representative uses are described in the "Immune Activity" and "infectious

"disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for

control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1038 of SEQ ID NO:23, b is an integer of 15 to 1052, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where b is greater than or equal to a + 14.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 14

The translation product of this gene shares sequence homology with chromaffin granule amine transporter protein which is thought to be important in vesicle membrane amine transport, particularly in the neural and endocrine tissue, and the human vesicular monoamine transporter hVMAT1 which is involved in the regulation of amine storage in cardiovascular, endocrine, and central nervous system function (See, Genbank Accession Nos. gill314290 and gblAAC50472.11; all references available through these accession numbers are hereby incorporated by reference herein). Based on these sequence similarities, The translation product of this gene is expected to share at least some biological activities with amine transporter proteins. Such activities are known in the art, some of which are described in Erickson, et al., PNAS 93:5166-5171 (1996), and/or Liu, et al., Cell 70:539-551 (1992), which are both incorporated herein by reference.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSFLDPTLSLFVLEKFNLPAQYVGLVFLGMAL
5 SYAISSPLFGLLSDKRPPLRKWLLVFGNLITAGCYMLLGPVPILHIKSQWLWLL
VLILVVSGLSAGMSIPTFPEILSCAHENGFEGLSTLGLVSGLFSAMWSIGAF
MGPTLGGFLYEKIGFEWAAAIQGLWALISGLAMGLFYLLYESRRKRKSKSQNIL
STEEERTTLLPNET (SEQ ID NO: 278). Polynucleotides encoding these
polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in colon cancer, osteoclastoma, and T-cell lymphoma, and to a lesser extent in many tumor or proliferative tissues such as
15 endometrial tumor, chondrosarcoma, induced umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases resulting from disorders in small molecule transport (i.e.,
20 signalling molecules) in afflicted tissues and organs, particularly of the endocrine and central nervous systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal, immune, and/or digestive systems
25 and cancer expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression
30 level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 137 as residues: Ser-114 to Asn-123, Thr-127 to Thr-132. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in colon cancer, osteoclastoma, and T-cell lymphoma and homology to amine transporter family members indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders or diseases resulted from small molecule transport in afflicted tissues and organs, particularly that of colon, osteoclast or T-cells. The expression in cancer tissues, and shared homology with transporter proteins may also indicate its role in anti-cancer drug resistance. Additionally, the protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; or for identifying inhibitors or promoters of the transport of toxic molecules to vesicles, for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1527 of SEQ ID NO:24, b is an integer of 15 to 1541, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with the human prolyl 4-hydroxylase alpha (II) subunit which is important in catalyzing the formation of 4-hydroxyproline in collagens which is essential for the folding of newly synthesised collagen polypeptide chains into triple-helical molecules (See Genbank Accession No. gblAAB71339.11; all references available through this accession are hereby incorporated herein by reference). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Prolyl 4-hydroxylase proteins. Such activities are known in the art, some of which are described in Annunen, et al., J. Biol. Chem. 272:17342-17348 (1997) which is incorporated herein by reference.

When tested against U937 myeloid and Jurkat T-cell cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates myeloid cells and T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence:

GTREARLRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSL
 EASENIRALKDGYEKVEQDLPAFEDLEGAARALMRLQDVYMLNVKGLAR
 GVFQRTGSAITDLYSPKRLFSLTGDDCFQVGKVAYDMGDYYHAIPWLEEA
 VSLFRGSYGEWKTEDEASLEDALDHLAFAYFRAGNVSCALSLSREFLLYSPD
 5 NKRMAENVLKYEYRLLAESPNHVVAEAVIQRPNIPHLQTRDTYEGLCQTL
 GSQPTLYQIPSLYCSYETNSNAYLLLQPIRKEVIHLEPYIALYHDFVSDSEAQ
 KIRELAEPWLQRSVVASGEKQLQVEYRISKSAWLKDTVDLKLVTLNHRIA
 LTGLDVRPPYAEYLQVVNYGIGGHYEPHFDHATSPSSPLYRMKSGNRVATFM
 IYLSSVEAGGATAFIYANLSVPVVRNAALFWWNLHRSGEGDSDTLHAGCP
 10 VLVGDKWVANKWIHEYGQEFRPCSSSPED (SEQ ID NO: 282). Additional.

Preferred polypeptides comprise the following amino acid sequence: GTREA
 RLRLDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSL EASENI
 RALKDGYEKVEQ DLPAFEDLEGAARAL (SEQ ID NO: 279), ALMRLQD (SEQ
 ID NO: 280), and/or VEAGGAT (SEQ ID NO: 281). Polynucleotides
 15 encoding these polypeptides are also provided.

This gene is expressed primarily in lymph node breast cancer, colon
 carcinoma, and to a lesser extent in osteoblasts and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions which include, but are
 not limited to, disorders of connective and immune tissues, particularly autoimmune
 disorders. Similarly, polypeptides and antibodies directed to these polypeptides are
 useful in providing immunological probes for differential identification of the
 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
 25 particularly of the connective tissues in breast, colon, bone, and fat, expression of this
 gene at significantly higher or lower levels is routinely detected in certain tissues or
 cell types (e.g., immune, connective, or cancerous and wounded tissues) or bodily
 fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another
 tissue or cell sample taken from an individual having such a disorder, relative to the
 30 standard gene expression level, i.e., the expression level in healthy tissue or bodily
 fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 138 as residues: Ser-74 to Ala-84, Gln-156 to Tyr-161, Tyr-184 to Asn-189, Ser-218 to Ile-223, Pro-299 to Ser-308, His-359 to Thr-368, Tyr-390 to Asp-404. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in lymph node breast cancer and colon carcinoma and homology to prol 4-hydroxylase alpha (II) subunit indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of connective tissue disorders and diseases (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), as well as, in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias ie. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Alternatively, the tissue distribution within various tissue carcinomas and tumor tissues, and biological activity reflected by the binding and activation of the GAS promoter element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders.

20 Expression in cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2065 of SEQ ID NO:25, b is an

30

integer of 15 to 2079, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

5 In an additional embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: IQPSHAALLHCRSTFRKTECLDPW
WVRRQLLG MAGIGGLQKMKAPHTGVLHLG SVWVFLGPFL LGVGYTLTFNPL
10 SGCMSTVRWLNSNITANRTLRSVCHVTPLHRSLSPHDGEYLRQMLLNSSSR
AGEAGSWG Y (SEQ ID NO: 283). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 20. Accordingly, polynucleotides related to this invention are useful as a marker in
15 linkage analysis for chromosome 20.

This gene is expressed primarily in fetal liver, and, to a lesser extent, in a variety of fetal and other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, liver disorders and cancers (e.g., hepatoblastoma, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hepatic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, bile, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an
30 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 139 as residues: Ser-67 to Tyr-75. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

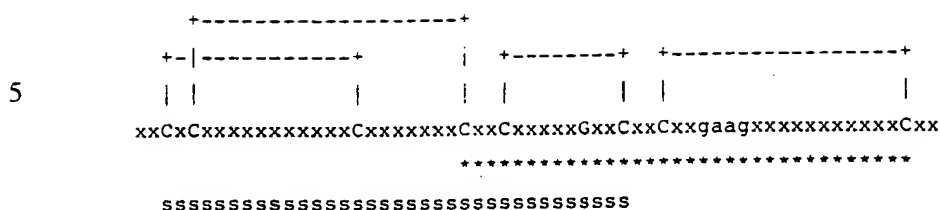
Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1933 of SEQ ID NO:26, b is an integer of 15 to 1947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this gene shares sequence homology with human laminin B1 which is thought to be an important structural extracellular matrix component involved in cell migration and signalling, particularly in stimulating epithelial cell growth and differentiation (See, Genbank Accession No gill186837).

Preferred polypeptides of the invention comprise the following amino acid sequences: CSSPPGRLPWCWTAPRTLKGHGSLLSTLRLTAPLHLAWKMMLS RKALFVLLNTPVLFHALEGRLEFSKLCCHHTIQRTLTVPKFRSS (SEQ ID NO: 284), RSPTSRVQLLKRQSCPCQRNDLNEEPQHFTHYAIYDFIVKGSCFCNG
 5 HADQCIPVHGFRPVKAPGTFHMHGKCM (SEQ ID NO: 285), and/or HNTAG SHCQHCAPLYNDRPWEAADGKTGAPNECRTCKCNGHADTCHFDVNVWEAS GNRSGGVCDDCQHNTGQYQCRCKPGFYRDLRRPFSAPDACKPCSCHPV GSAVLPAHSVTFCDPSNGDCPCCKPGVAGRRCDRCMVGYWGFGDYGCRP CDCAGSCDPITGDCISSHTDIDWYHEVPDFRPVHNKSEPAWEWEDAQGFSAL
 10 LHSKGCECKEQLGNAKAFCGMKYSYVLKIKILSAHDKGTHVEVNVKIK KVLKSTKLKIFRGKANIISRIMDGQ RMHLSNPQSWFGIPCSRT (SEQ ID NO: 286). Polynucleotides encoding these polypeptides are also provided.

Included in this invention as preferred domains are Laminin-type EGF-like (LE) domain signatures, which were identified using the ProSite analysis tool (Swiss
 15 Institute of Bioinformatics). Laminins are the major noncollagenous components of basement membranes that mediate cell adhesion, growth migration, and differentiation. They are composed of distinct but related alpha, beta and gamma chains. The three chains form a cross-shaped molecule that consist of a long arm and three short globular arms. The long arm consists of a coiled coil structure contributed
 20 by all three chains and cross-linked by interchain disulfide bonds. Beside different types of globular domains each subunit contains, in its first half, consecutive repeats of about 60 amino acids in length that include eight conserved cysteines. The tertiary structure of this domain is remotely similar in its N-terminal to that of the EGF-like module. It is known as a 'LE' or 'laminin-type EGF-like' domain. The number of
 25 copies of the LE domain in the different forms of laminins is highly variable: from 3 up to 22 copies have been found. A schematic representation of the topology of the four disulfide bonds in the LE domain is shown below.



10

'C': conserved cysteine involved in a disulfide bond

'a': conserved aromatic residue

'G': conserved glycine (lower case = less conserved)

15 's': region similar to the EGF-like domain

'*': position of the pattern

In mouse laminin gamma-1 chain, the seventh LE domain has been shown to be the only one that binds with a high affinity to nidogen. The binding-sites are

20 located on the surface within the loops C1-C3 and C5-C6. Long consecutive arrays of LE domains in laminins form rod-like elements of limited flexibility, which determine the spacing in the formation of laminin networks of basement membranes. We derived a signature pattern for the LE domain which covers the C-terminal half of the repeat starting with the fourth conserved cysteine. The consensus pattern is as

25 follows: C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C [All C's are involved in disulfide bonds]

Preferred polypeptides of the invention comprise the following amino acid sequence: CDDCQHNTGQYCQRCKPGFYRDLRRPFSAPDACKPC (SEQ ID NO: 287) and/or CPCKPGVAGRRCDRCMVGYWGFGDYGCRPCDCAGSC (SEQ

30 ID NO: 288). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the laminin-type EGF-like domains listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence encoded by this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the laminin-type EGF-

like domain. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the laminin-type EGF-like domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domain is characteristic of a signature specific to

- 5 Laminin proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Laminin proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in osteoblastic tissues and cell types, including osteoblasts, osteoblastomas and osteoclastomas. Expression is also
10 abundant in vascular-pulmonary tissues such as lung, micro-vasculature, pulmonary, endothelial and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
15 not limited to, cancer and malignancies (particularly of osteoblastic tissues and rhabdomyosarcoma), as well as cardiovascular and respiratory or pulmonary disorders such as athsma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stoke, thrombosis hypertension, inflammation and wound healing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological
20 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardio-respiratory system, and skeletal system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, osteoblast, cardio-respiratory, vascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,
25 pulmonary surfactant, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
30 epitopes shown in SEQ ID NO: 140 as residues: Ser-28 to Cys-34, Thr-51 to Thr-58, Tyr-64 to Asn-81, Asp-111 to Lys-116, Asp-145 to Phe-160, Pro-203 to Glu-217. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in osteoblastic tissues and cell types and homology to laminin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and diagnosis of cardiovascular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis hypertension, inflammation and wound healing. As a homolog of laminin, this gene product quite possibly has a role in cell adhesion, migration, proliferation, angiogenesis, chondrogenesis, wound healing and oncogenesis. An EST (Int J Cancer 1996 May 16;66(4):571-577) with an identical sequence to part of this contig was shown to be differentially expressed in human primary myoblasts and embryonal rhabdomyosarcoma and therefore might have an important role in the determination or maintenance of the normal phenotype, and thus its loss is possibly involved in the progression of malignancies, particularly of skeletal muscle. Similarly, the homology to a laminin would suggest a role in the detection and treatment of disorders and conditions afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation) in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 3365 of SEQ ID NO:27, b is an integer of 15 to 3379, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The gene encoding the disclosed cDNA is believed to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

10 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: NISSQYCILKSLEMMISGLKLLVLFLKFAPENY CLSTETLQMPNRHLRLSKATCYLMKCLLPSEYFE (SEQ ID NO: 289).

Polynucleotides encoding these polypeptides are also provided.

15 This gene is expressed primarily in placenta, brain, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, behavioral, or nervous system disorders, such as: depression, schizophrenia, Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, dementia, paranoia, addictive behavior, epilepsy, transmissible spongiform encephalopathy (TSE), Creutzfeldt-Jakob disease (CJD). Other diseases and conditions related to expression in the placenta might include developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive disfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and reproductive systems, 25 expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial 30

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 141 as residues: Ala-16 to Leu-22. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, expression in placenta would suggest a possible role in the treatment and diagnosis of developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders.

Similarly, expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may

show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1992 of SEQ ID NO:28, b is an integer of 15 to 2006, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

The translation product of this gene shares sequence homology with the murine transforming protein (See, e.g., Genbank Accession No. gil53529|embl|CAA36859.1|; all references available through this accession are hereby incorporated by reference herein).

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PIEGTPAGTGPEFPGRPTRPQRMRS LISSHPCQ HLLLLLLLLFLILAILVDVKWYLVLFICISLMTSDVEHLFMCLLAIRISSWR NVY (SEQ ID NO: 290). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in activated and basal T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, auto-immunities, cancer, metastasis, wound healing, inflammation, anemias (leukemia) and other hematopoietic disorders. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 10 The tissue distribution in activated T-cells and the homology to a murine transforming protein indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere
- 15 herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).
- 20 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,
- 25 such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted
- 30 factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of

various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials: for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3056 of SEQ ID NO:29, b is an

integer of 15 to 3070, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

5 The translation product of this gene was shown to have homology to the Mus musculus ALG-2 protein, which is known to code for a Ca(2+)-binding protein required for T cell receptor-, Fas-, and glucocorticoid-induced cell death. ALG-2 mediate Ca(2+)-regulated signals along the death pathway and may play a role in the onset of Alzheimer's Disease (See e.g., Genbank Accession No.gil1213520; all
10 references available through this accession are hereby incorporated by reference herein).

 Preferred polypeptides comprise the following amino acid sequence: NWVPT CLCP SAPCSFHLLSRFKCLFSPQRLTDIFRRYDTDQDGWIQVSYEQYLSMVFS
15 IV (SEQ ID NO: 291), and/or QRLTDIFRRYDTDQDGWIQVSYEQYLSMVFSIV (SEQ ID NO: 292). Polynucleotides encoding these polypeptides are also provided.

 When tested against K562 cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element). Thus, it is likely that this gene activates immune or leukemia cells through the Jaks-STAT signal transduction pathway. ISRE is a promoter element found upstream in
20 many genes which are involved in the Jaks-STAT pathway. The Jaks-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

25 A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MFYKLT LILCELSVAGVTQAASQRPLQRLPRHICSQR XPPGRCLLKAXLQTTWXXPKPI PRLSPPLXSDPKR (SEQ ID NO: 293). Polynucleotides encoding these polypeptides are also provided.

 The gene encoding the disclosed cDNA is believed to reside on chromosome
30 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

This gene is expressed primarily in placenta, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 143 as residues: Arg-24 to Arg-31, Ile-33 to Gly-41. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and/or diagnosis of developmental anomalies, fetal deficiencies, ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells combined with the observed ISRE activity, and homology to the apoptosis linked, ALG-2 indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:30 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2213 of SEQ ID NO:30, b is an integer of 15 to 2227, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:30, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

The translation product of this gene was shown to have homology to the human histo-blood group A transferase (See, e.g., Genbank Accession No. gblAAD26573.1|AF134413_1 (AF134413); all references available through this
 5 accession are hereby incorporated by reference herein) which is known to represent one of the major allogeneic antigens in both erythrocytes and tissues of humans. Its been proposed that the A phenotype is associated with the glycosyltransferase that converts the H substance associated with the O phenotype to A through the addition of alpha 1-3-N-acetylgalactosamine or alpha 1-3-galactosyl residues to the H antigen
 10 Fuc-alpha 1-2Gal- beta 1-R. Therefore, the primary product of the histo-blood group A is its respective glycosyltransferase. Preferred polypeptides of the invention comprise the following amino acid sequence: TSSPVFSFCSMAVREPDHLQ
 RVSLPRYNVSASLQWLPCHRIVLQPWHMCAMWELGQVLFHPVAPREGAAPS
 PVSTLTWPSSCSHSESTMELELQF (SEQ ID NO: 294), LPCHRIV (SEQ ID NO:
 15 296), SLQWLPCHRIVLQPW (SEQ ID NO: 297), and/or MAVREPDHLQRVSLPR (SEQ ID NO: 295). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in 12-week-old human embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, pre-natal disorders, hematopoietic disorders, or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
 25 tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic, lymph, developing, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder.
 30 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in 12 week old embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental anomalies, fetal deficiencies, pre-natal disorders and cancer. Similarly, expression within embryonic tissue and other cellular sources marked by

5 proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Alternatively, the tissue distribution and homology to human blood group A

10 and B glycosyltransferase enzymes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious

15 disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be

20 used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to

25 isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly

30 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:31 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically

excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1274 of SEQ ID NO:31, b is an integer of 15 to 1288, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:31, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

The translation product of this gene shares sequence homology with CD97 (EMR1), which is thought to be important in both adhesion and signaling processes early after leukocyte activation (See, e.g., Genbank Accession No. gil784994; all references available through this accession are hereby incorporated by reference herein). EMR1 belongs to a novel family of G-protein receptors that has recently been recognized on the basis of homologous primary amino acid sequences, comprises receptors to hormones of the secretin/vasoactive intestinal peptide/glucagon family, parathyroid hormone and parathyroid hormone-related peptides, growth hormone-releasing factor, corticotropin-releasing factor, and calcitonin. Proteins with seven transmembrane segments (7TM) define a superfamily of receptors (7TMreceptors) sharing the same topology: an extracellular N-terminus, three extramembranous loops on either side of the plasma membrane, and a cytoplasmic C-terminal tail. Upon ligand binding, cytoplasmic portions of the activated receptor interact with heterotrimeric G-coupled proteins to induce various second messengers, which subsequently activate various signal transduction pathways depending upon the specific G-coupled protein associated with the receptor. Preferred polypeptides of the invention comprise the following amino acid sequence: CFKRKPKREHCSCP
ITYQSLGDILNASFFSKRKGMQEVKLNSYVVS GTIGLKEKISLSEPVFLTFRHN
QPGDKRTKHICVY WEGSEGGRWSTEGCSHVHSNGSYTKCKCFHLSSFAVLV
ALAPKEDPVLTVITQVGLTISLLCLFLAILTFLLCRPIQNTSTSLHLELSLCLFLA
HLLFLTGINRTEPEVLCSIIAGLLHFLYLACFTWMLLEGLHLFLTVRNLKVAN
YTSTGRFKKRFMPVGYGIPAVIIAVSAIVGPQNYGTFTHCWLKLDKGFIWSF
MGPVAVIILINLVFYFQVLWILRSKLSSLNKEVSTIQDTRVMTFKAISQLFILGC
SWG LGFFMVEEVGKTIGSIIAYSFTIINTLQGVLLFVVHCLLNQRQVRMEYKKW

FSGMRKGVETESTEMSRSTTQTKTEEVGKSSEIFHKGGTASSSAESTKQPQPQ
 VHLVSAAWLKMN (SEQ ID NO: 298), and/or FFWKENLRRNGSREDFARRATQ
 LIQSVELSIWNASFASPGKGQISEFDIVYETKRCNETRENAFLEAGNNTMDINC
 ADALKGNLRESTAVALSLINLLGIF SEQ ID NO: 299. Polynucleotides

5 encoding these polypeptides are also provided.

Included in this invention as preferred domains are two EGF-like protein domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). First, a sequence of about forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown to be present in a large number of membrane-bound and extracellular, mostly animal proteins. Many of these proteins require calcium for their biological function and a calcium-binding site has been found to be located at the N-terminus of some EGF-like domains. Calcium-binding is crucial for numerous protein-protein interactions. We have used the N-terminal part of the EGF domain as a consensus pattern. It includes the negative N-terminus and the possible hydroxylation site. The consensus pattern is as follows: [DEQN].[DEQN]{2}C.{3,14}C.{3,7}C.[DN].{4}[FY].C [The four C's are involved in disulfide bonds].

Preferred polypeptides of the invention comprise the following amino acid sequence: DINECETGLAKCKYKAYCRNKVGGYIC (SEQ ID NO: 300). Polynucleotides encoding these polypeptides are also provided. Secondly, post-translational hydroxylation of aspartic acid or asparagine to form erythro-beta-hydroxyaspartic acid or erythro-beta-hydroxyasparagine has been identified in a number of proteins with domains homologous to (EGF). Based on sequence comparisons of the EGF-homology region that contains hydroxylated Asp or Asn, a consensus sequence located in the N-terminal of EGF-like domains has been identified that seems to be required by the hydroxylase(s). The consensus sequence is as follows: C.[DN].{4}[FY].C.C.

Preferred polypeptides of the invention comprise the following amino acid sequence: CRNKVGGYICSC (SEQ ID NO: 301). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the calcium-binding EGF-like domain and aspartic acid and asparagine hydroxylation site listed above, and at least

5. 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene and the embodiments listed herein. The additional contiguous amino acid residues is N-terminal or C-terminal to one or both of the listed domains. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to one or both of the listed domains, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to EGF like proteins. Based on the sequence similarity and conserved domains. The translation product of this gene is expected to share at least some biological activities with EGF-like proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in eosinophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders or anemias and leukemias, immunodeficiencies, infection, lymphomas, auto-immunities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 145 as residues: Ser-22 to Ser-30, Pro-33 to Cys-48, Asp-50 to Lys-67, Pro-117 to Ser-130. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in eosinophils combined with its homology to a known human seven transmembrane domain protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders, particularly considering the fact that the majority of 7 transmembrane receptors are tightly associated with signal transduction pathways which are integral to the modulation of the cell cycle. As such, the protein product of this gene may play a role in the regulation of cellular division, where loss of regulation may result in proliferating cells and the onset of tumors or cancer. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Similarly, the tissue distribution and homology to CD97 indicates that the protein product of this gene might be a marker for differentiation and activation of eosinophils, and therefore is useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g., AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:32 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
 5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3266 of SEQ ID NO:32, b is an integer of 15 to 3280, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:32, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

The translation product of this gene has been found to have homology to the rat neural F box protein NFB42, in addition to a conserved *Caenorhabditis elegans* C14B1.3 protein (See, e.g., Genbank Accession Nos. gil3851648|gb|AAC97505.11 (AF098301) and gil558270; all references available through these accessions are
 15 hereby incorporated by reference herein). Preferred polypeptides of the invention comprise the following amino acid sequence: ALCPPHLLNVTVSPAPSCRHVK KVVASPPSTTFMIAMDAPHSKAALDSINELPENILLELFTHVPARQLLLNCRL VCSLWRDLIDLMTLWKRKCLREGFITKDWDQPVADWKIFYFLRSLHRNLLR NPCAEEDMFAWQIDFNGGDRWKVESLPGAHGTDFFDPKVKKYFVTSYEMCL
 20 KSQVLVDLVAEGYWEELLDTRPDIVVKDWFAARADCGCTYQLKVQLASA DYFVLASFEPPTVTIQWNNATWTEVSYTFSDYPRGVRYLQFHGGRDTQY WAGWYGPRVTNSSIVVSPKMTRNQASSEAPGQKHGQEEAAQSPYRAVV QIF (SEQ ID NO: 302). Polynucleotides encoding these polypeptides are also provided.

25 The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune cells, especially primary dendritic cells, and T cells, and to a lesser extent in a variety of other tissues including breast,
 30 keratinocytes, epididymus (cauda), lung, multiple sclerosis, endometrial stromal cells, IL4 induced umbilical vein endothelial cells, fetal kidney, fetal dura mater, rejected kidney, and osteoblasts.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders, particularly of the immune system or endothelial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 146 as residues: Pro-41 to Cys-47, Phe-52 to Gly-59, Pro-62 to His-70. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in T-cells indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:33 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
5 formula of a-b, where a is any integer between 1 to 1283 of SEQ ID NO:33, b is an integer of 15 to 1297, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:33, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

10 The translation product of this gene shares sequence homology with the human, mouse, and bovine dopamine hydroxylase which is thought to be important in the modification of dopamine, a neurotransmitter (See Genbank Accession Nos. gil30474, gil162965, and/or gil2358082; all references available through these accessions are hereby incorporated by reference herein). Preferred polypeptides of the
15 invention comprise the following amino acid sequence: RQRSWNPQT
NCYHPNMPDAFLTCETVIFAWAIGGEGFSYPPHVGLSLGTPLDPHYVLLEVH
YDNPTYEEGLIDNSGLRLFYTMDIRKYDAGVIEAGLWVSLFHTIPPGMPEF
QSEGHCTLECLEEALAEKPSGIHVFAVLLHAHLAGRGIRLRHFRKGKEMKL
LAYDDDFDFNFQEFQYLKEEQTILPGDNLITECRYNTKDRAEMTWGGLSTR
20 SEMCLSYLLYYPRINLTRCASIPDIMEQLQFIGVKEIYRPVTTWPFIIKSPKQYK
NLSFMDAMNKFkwTKKEGLSFNKLVLSPVNVRCskTDNAEWSIPRNDsIT
SRYRKTL (SEQ ID NO: 303). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following
25 amino acid sequence: MCCWPLLLLWGLLPgTAAGGSGRTYPHRTLLDSEGK
YWLGWSQRGSQIAFRLQVrTAGYVGFGFSPTGAMASADIVVGGVAHGR
PYLQDYFTNANRELKKDAQQDYHLEYAMENSTHTIIEFTRELHTCDINDKS
ITDSTVRVIWAYHHE DAGEAGPKYHDSNRGtKSLRLLNPEKTSVLSTALPYF
DLVNQDVPIPNKDTTYWCQMFKIPVFQEKHHVIKVEPVIQRGHESLVHHILL
30 YQCSNNFNDSVPGIRARIAITPTCPMHSSPV KL (SEQ ID NO: 304).
Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in brain, the pulmonary system, and to a lesser extent in kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological and behavioral disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., sputum, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 147 as residues: Ser-33 to Trp-38, Gly-40 to Gly-45, Asn-93 to Asp-105, Thr-128 to Thr-137, Glu-150 to Lys-167, Pro-197 to Tyr-203, Cys-242 to Asn-247, Ser-253 to Tyr-258, His-307 to Glu-314, Glu-357 to Gly-362, Trp-373 to Gln-378, Ser-402 to Glu-408. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to a protein involved in the modification of dopamine indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive

compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Alternatively, the homology to dopamine hydroxylase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:34 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2170 of SEQ ID NO:34, b is an integer of 15 to 2184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:34, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway.

The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that
5 this gene activates T-cells through the JakStat signal transduction pathway.

This gene is expressed in a variety of human normal and diseased tissues including breast, infant adrenal gland, skin tumor, colon, pituitary, Wilm's tumor, and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer and other proliferative disorders, afflicting endocrine or endothelial tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system or of breast and/or breast lymph nodes, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial
20 fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast, infant adrenal gland, skin tumor, colon, pituitary, and Wilm's tumor, and biological activity in activating the GAS promoter
25 element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-,
30 hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Alternatively, the tissue distribution and biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the

diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells. (i.e., breast, skin and Wilm's tumors) indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues
5 indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and
10 differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine
15 biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

20 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:35 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
25 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 935 of SEQ ID NO:35, b is an integer of 15 to 949, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:35, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the
5 following amino acid sequence: TGTFWSPRSQRRGCCGRRAPRPEAMENGAVYS
PTTEEDPGPARGPRSGLAAYFFMGRLPLLRVLKGLQLLSLLAFICEEVVSQ
CTLCGGLYFFEFVSCSAFLLSLLILIVYCTPFYERVDTTKVKSSDFYITLGTGCV
FLLASIIFVSTHDRTSAEIAAIVFGFIASFMLLDFITMLYEKRQESQLRKPENTT
RAEALTEPLNA (SEQ ID NO: 305). Polynucleotides encoding these polypeptides
10 are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 3. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 3.

This gene is expressed primarily in dendritic cells, and to a lesser extent in
15 melanocytes, fetal liver and spleen and several other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammation, and disorders of the hepatic and immune systems.
20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic,
25 hepatic, immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 149 as residues: Phe-63 to Ser-75, Thr-97 to Ser-102, Glu-128 to Arg-143. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in dendritic cells indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, raise

antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

5 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:36 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
10 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3324 of SEQ ID NO:36, b is an integer of 15 to 3338, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:36, and where b is greater than or equal to a + 14.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the
20 following amino acid sequence: ASAPRVMRGHLAGFPALSGLASVCLWATFSA QLPGPVAATSWTPAPLGCSAARSQPEKRLGTAAPGSAASLAQAGPGAPCRV LPVDPAPAALNVREPGWLGGFLDGALLQVLLNFLRKSTDVLMDTREAESLEV E (SEQ ID NO: 306).

In another embodiment polypeptides of the invention comprise the following
25 amino acid sequence: NKLHSFPVFLSQLLLDRQLLHAPQTLPTPHCGGSSRPGP SHPPWLLIQLPCVHVALWQMLRDFSDSRITPSTLTTQPAAQTAAPAKDQES DIVGGEGILCDIAFLQEDHPLGVGGASAPSSRRELSRRGVHTQTLPEDGTLHG TPSSSFDCGIKYIISWPLAPGCDLPSLELSLVCKGVSSCMGFAAG (SEQ ID NO: 307). Polynucleotides encoding these polypeptides are also provided.

30 This gene is expressed primarily in endothelial cells, lung, and fetal kidney, and to a lesser extent in epididymis, keratinocytes and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cardiovascular diseases involving endothelial cell disturbances such as atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 150 as residues: Arg-47 to Leu-54. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in endothelial cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating disorders of endothelial cells such as atherosclerosis, vasculitis, cardiovascular disease, and emphysema. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. The polypeptide may possess a wide range of undetected biological activities. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or

thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the
5 corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
10 related to SEQ ID NO:37 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1549 of SEQ ID NO:37, b is an integer of 15 to 1563, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:37, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

20 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PGRPTRPTKNKVCVCLGMLFWAYPICVFIDSL
SCQPCLWSTGATSHFNSPTTSPLFTLFMPICALAPNPFT QLGLDDDR (SEQ ID
25 NO: 308). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
30 not limited to, tumors or disorders of the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 151 as residues: His-29 to Thr-34. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in meningioma indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception, as well as disorders of the meninges such as meningioma and meningitis. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:38 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1034 of SEQ ID NO:38, b is an integer of 15 to 1048, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:38, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

The translation product of this gene has been shown to encode a human brain specific mitochondrial carrier (Genbank Accession No. gil3851540|gb|AAD04346.11 (AF078544); all references available through this accession are hereby incorporated herein by reference) which shares sequence homology with the human body weight disorder associated gene C5 product which is known to be differentially expressed in obese compared to lean mice (See GeneSeq Accession No. R91281). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with mitochondrial carriers proteins. Such activities are known in the art, some of which are described in Sanchis et al. J. Biol. Chem. 273:34611-34615 (1998), incorporated herein by reference.

Included in this invention as preferred domains are mitochondrial energy transfer protein (METP) domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Structurally, members of the family of mitochondrial energy transfer proteins consist of three tandem repeats of a domain of approximately one hundred residues. Each of these domains contains two transmembrane regions. As a signature pattern, we selected one of the most conserved regions in the repeated domain, located just after the first transmembrane region. To detect this widespread family of proteins, a consensus sequence was developed that contains the most conserved regions in the repeated domain. The consensus pattern is as follows: P.[DE].[LIVAT][RK].[LRH][LIVMFY][QMAIGV].

Preferred polypeptides of the invention comprise the following amino acid sequences: PVDLTKTRLQ (SEQ ID NO: 309) and PTDVLKIRMQ (SEQ ID NO: 310). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the METP domains of the
5 sequence listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional
contiguous amino acid residues of the sequence referenced in Table I for this gene.
The additional contiguous amino acid residues is N-terminal or C-terminal to the
METP domain. Alternatively, the additional contiguous amino acid residues is both
N-terminal and C-terminal to the METP domain, wherein the total N- and C-terminal
10 contiguous amino acid residues equal the specified number. The above preferred
polypeptide domain is characteristic of a signature specific to mitochondrial energy
transfer proteins.

The gene encoding the disclosed cDNA is believed to reside on chromosome
X. Accordingly, polynucleotides related to this invention are useful as a marker in
15 linkage analysis for chromosome X.

This gene is expressed primarily in brain, and to a lesser extent, in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
20 not limited to, neurological and behavioral disorders and immune disorders and/or
obesity. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the
tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
particularly of the digestive, immune, and nervous systems, expression of this gene
25 at significantly higher or lower levels is routinely detected in certain tissues or cell
types (e.g., immune, neural, or cancerous and wounded tissues) or bodily fluids (e.g.,
lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell
sample taken from an individual having such a disorder, relative to the standard gene
expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 152 as residues: Gln-189 to Gly-195. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to mitochondrial carrier proteins indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:39 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1416 of SEQ ID NO:39, b is an integer of 15 to 1430, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:39, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

Preferred polypeptides of the invention comprise the following amino acid sequence: MTFGSTISPTSTHASPGLGCCSWLLEDLEEQLYCSAFEEAALTR
RICNPTSCWLPLDMELLHRQVLALQTQRVLLGMWLRRAWDTWVSPRRVAP
GSRCLLTASHPCTEKRRKASAXQRNLGYPLAMLCLLVLTGLSVLIVAIHILEL
10 LIDEAAMPGRMQGTSLGQVSFSKLGSGAVIQVVLIFYLMVSSVVGIFYSSPLF
RSLRPRWHDTAMTQIIGNCVCLLVSSALPVFSRTLGLTRFDLLGDFGRFNWL
GNFYIVFLYNAAFAGLTTCLVKTFATAVRAELIRAFGLDRLPLPVSGFPQAS
RKTQHQ (SEQ ID NO: 311). Polynucleotides encoding such polypeptides are also provided.

15 This gene is expressed primarily in immune system tissues (e.g. resting T-cells, primary dendritic cells, and neutrophils, apoptotic T-cells) and umbilical vein. This gene is expressed to a lesser extent in the gastrointestinal tissue (e.g. small intestine, colon), brain (e.g. cerebellum, frontal cortex), aorta endothelial cells, skin tumor, embryonic tissue, thymus, and cancers (e.g. cheek, breast, synovial).

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for
25 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and gastrointestinal tract expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and
30 spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 153 as residues: Asp-21 to Ser-29. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells (e.g. T-cells, dendritic cells, 5 neutrophils) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the 10 proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene 15 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and 20 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits 25 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in skin tumors and cancerous tissue (e.g. cheek, breast, synovial sarcoma) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the 30 diagnosis and treatment of cancer and other proliferative disorders. Expression in cellular sources such as embryonic tissue marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the

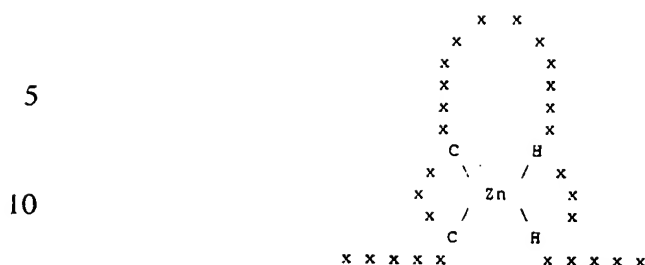
expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from
5 early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or
10 immunotherapy targets for the above listed tissues. The tissue distribution in cerebellum and frontal cortex indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative
15 Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction,
20 aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.
25 Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional
30 supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:40 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2089 of SEQ ID NO:40, b is an integer of 15 to 2103, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:40, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The polypeptide of this gene has been determined to have a zinc finger (Zinc finger, C2H2 type) domain at about amino acid position 16-50 of the amino acid sequence referenced in Table I for this gene. Therefore,

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: LCVCLVYLCMYGVCLCVIVCVSGVSLCLYVWGV SVC DCVSVFMCVCLCVIFCVYGKPRTEHYHSPHLAKQKAFREMCGRHDVSAAGIF QSYV (SEQ ID NO: 312). Polynucleotides encoding these polypeptides are also provided. 'Zinc finger' domains are nucleic acid-binding protein structures first identified in the *Xenopus* transcription factor TFIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:



Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc- dependent DNA or RNA binding property of some members of this class. Some of the proteins known to include C2H2-type zinc fingers are listed below. We have indicated, between brackets, the number of zinc finger regions found in each of these proteins; a '+' symbol indicates that only partial sequence data is available and that additional finger domains is present.

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, salivary gland related diseases, diseases of the mouth, and other digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., saliva, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 154 as residues: Gly-46 to His-54. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that the protein products of this gene are
5 useful for diagnosis and treatment of salivary gland related diseases (mumps, calculi formation in ducts, sarcoidosis, facial palsy, tumors, Sjogrens Syndrome) and other digestive system disorders. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a
10 nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
15 related to SEQ ID NO:41 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
20 formula of a-b, where a is any integer between 1 to 2335 of SEQ ID NO:41. b is an integer of 15 to 2349, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:41, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 32

25 This gene is expressed primarily in fetal tissue (e.g. spleen, liver, brain), cancerous tissues (e.g. ovarian, colon, stomach, parathyroid) and to a lesser extent in immune cells and tissue (e.g. B-cells, T-cells, bone marrow), and reproductive organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly of the colon and ovaries, disorders of the developing fetus, neurodegenerative conditions, and immune system disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is
5 routinely detected in certain tissues or cell types (e.g., immune, reproductive, neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 155 as residues: Lys-35 to Lys-47. Polynucleotides encoding said polypeptides are also provided.

The expression of this gene within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular
15 division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern
20 formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of
25 potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types
30 of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in

modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (such as T-cells and B-cells) and immune tissues (bone marrow) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases; or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in parathyroid indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for

the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Additionally, the tissue distribution in brain tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:42 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1545 of SEQ ID NO:42, b is an integer of 15 to 1559, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:42, and where b is greater than or equal to a + 14.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma
10 activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and
15 differentiation of cells.

This gene is expressed primarily in skin tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
20 not limited to, skin disorders, particularly skin cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell
25 types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 156 as residues: Pro-38 to Gly-44, Phe-56 to Thr-64. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in skin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders including congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. 5 keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, 10 pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. Moreover, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athletes foot, and ringworm). Protein, as well as, antibodies directed against the 15 protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:43 and may have been publicly available prior to conception of 20 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1752 of SEQ ID NO:43, b is an 25 integer of 15 to 1766, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:43, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

The translation product of this gene shares sequence homology with mitogen- 30 induced prostate carcinoma (mouse) which is thought to be important in the etiology of cancer. In this respect, this gene is mitogen-induced and/or involved in cell proliferation.

Preferred polypeptides of the invention comprise the following amino acid sequence: GHMPYGWLTEIRAVYPAFDKNNPSNKLVSTSNTVTAHHKKF TFVCMALSLTLCFVMFWTPNVSEKILIDIIGVDFAFELCVVPLRIFSFFPVPVT VRAHLTGWLMTLKKTFVLAPSSVLRIIVLIASLVVLPYLGVBHGATLGVGSLLA
5 GFVGESTMVAIAACYVYRKQKKK MENESATEGEDSAMTDMPTTEEVTDIVE MRENE (SEQ ID NO: 313) and/or QVVFVAILLHSHLECREPLLIPLSLYMGA LVRCTTLCLGYYKNIHDIIPDRSGPELGGDATIRKMLSFWWPLALILATQRRIS PIVNLFVSRDLGGSSAATEAVAILTATYPV (SEQ ID NO: 314). Polynucleotides encoding these polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

This gene is expressed primarily in early infant and adult brain, retina, fetal tissue (e.g., liver, spleen, whole embryo) and to a lesser extent in immune cells (e.g.,
15 monocytes and T-cells), colon, and parathyroid tumor tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, disorders of the immune system and nervous system.
20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic system (cancers), expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, neural,
25 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
30 epitopes shown in SEQ ID NO: 157 as residues: Arg-122 to Ser-139, Met-144 to Glu-149. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution and homology to mitogen induced prostate carcinoma (mouse) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study and treatment of cancers, including but not limited to the colon, parathyroid, and adrenal glands. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (T-cells, monocytes) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or

activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

- 5 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that
- 10 influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are
- 15 useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease,
- 20 Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism,
- 25 and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.
- 30

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to
5 identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
10 related to SEQ ID NO:44 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
15 formula of a-b, where a is any integer between 1 to 2558 of SEQ ID NO:44, b is an integer of 15 to 2572, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:44, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

20 This gene is expressed primarily in adult pulmonary tissue, umbilical vein, prostate, and fetal tissue (e.g., heart).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
25 not limited to, diseases of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in
30 certain tissues or cell types (e.g., pulmonary, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 158 as residues: Arg-45 to Gly-51, Glu-75 to Asn-81.

5 Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in pulmonary tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of disorders associated with developing lungs, particularly in premature infants where the lungs are the last tissues to develop. Additionally, the tissue distribution indicates
10 that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and intervention of lung tumors, since the gene is involved in the regulation of cell division, particularly since it is expressed in fetal tissue. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may
15 show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

20 Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have
25 applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue
30 differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in

proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:45 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 512 of SEQ ID NO:45, b is an integer of 15 to 526, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:45, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene is expressed primarily in adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fat metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 159 as residues: Pro-96 to Ser-106. Polynucleotides
5 encoding said polypeptides are also provided.

The tissue distribution in adipose tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of obesity and other metabolic and endocrine conditions or disorders. Furthermore, the protein product of this gene may show utility in ameliorating conditions which occur
10 secondary to aberrant fatty-acid metabolism (e.g. aberrant myelin sheath development), either directly or indirectly. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:46 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1018 of SEQ ID NO:46, b is an integer of 15 to 1032, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:46, and where b is greater than or equal to a + 14.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in adult brain tissue, testes, placenta, kidney, infant and fetal tissue (e.g., liver, spleen, lung) and to a lesser extent in immune cells (e.g., T-cells and neutrophils) and in cancerous tissues (e.g., ovarian tumor, Hodgkins lymphoma, pancreas, T-cell).

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, CNS disorders, disorders of the testicles, cancer, particularly ovarian, pancreatic, T-cell, and Hodgekin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS, and testes expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, urogenital, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders"

and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Additionally, the tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:47 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2666 of SEQ ID NO:47, b is an integer of 15 to 2680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:47, and where b is greater than or equal to a + 14.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

When tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates fibroblast cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation.

This gene is expressed primarily in endometrial stromal cells, endometrial tumors, keratinocytes, fetal tissue (e.g. liver, spleen) and to a lesser extent in endothelial cells and immune cells (e.g., T-cells).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial carcinoma and immune cells disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be

transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelihood of implantation and development of a healthy embryo. In both cases this gene or its gene

5 product could be administered at later stages of pregnancy to promote healthy development of the endometrium. Additionally, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endometrial carcinoma. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and
10 other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in
15 the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again
20 be useful in cancer therapy. The tissue distribution in immune cells such as helper T-cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders.

Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein.

25 Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

30 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such

as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and

chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:48 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1716 of SEQ ID NO:48, b is an integer of 15 to 1730, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:48, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in LNCAP cells (prostate cell line) and retina derived N2b5HR cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate cancer and eye disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urogenital and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative

to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 162 as residues: Asn-50 to Ser-57. Polynucleotides
5 encoding said polypeptides are also provided.

The expression in prostate may indicate the gene or its products can be used in the disorders of the prostate, including inflammatory disorders, such as chronic prostatitis, granulomatous prostatitis and malacoplakia, prostatic hyperplasia and prostate neoplastic disorders, including adenocarcinoma, transitional cell carcinomas,
10 ductal carcinomas, squamous cell carcinomas, or as hormones or factors with systemic or reproductive functions. The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans,
15 and retinoblastoma, retinochoroiditis, retinopathy and retinoschisis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or
20 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:49 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically
25 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1261 of SEQ ID NO:49, b is an integer of 15 to 1275, where both a and b correspond to the positions of nucleotide
30 residues shown in SEQ ID NO:49, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the

5 following amino acid sequence: RCCCRGCSCRARLCPPARSTAVAPECRGAHPSR
AMRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRRGTQRPCYKVIYFHD
TSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID NO: 315).

Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following
10 amino acid sequence: MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRRG
TQRPCYKVIYFHDTSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID
NO: 316). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle and human thyroid and to a
lesser extent in amniotic cells and human endometrial stromal cells-treated with
15 progesterone.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, thyroid disorders. Similarly, polypeptides and antibodies directed to
20 these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the endocrine system, expression of this gene at
significantly higher or lower levels is routinely detected in certain tissues or cell types
(e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine,
25 synovial fluid and spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

Preferred polypeptides of the present invention comprise immunogenic
30 epitopes shown in SEQ ID NO: 163 as residues: Ser-75 to Leu-81. Polynucleotides
encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endocrine disorders of the thyroid.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:50 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1748 of SEQ ID NO:50, b is an integer of 15 to 1762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:50, and where b is greater than or equal to a + 14.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 41**

This gene is expressed primarily in human testes tumor and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the testicles including but not limited to testicular cancer and immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and immune system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 164 as residues: His-31 to Gly-41. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in testes, particularly testicular tumors, indicates that

5 polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists)

10 are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as

15 hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. The tissue distribution in bone marrow indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in

20 Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the

25 treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory

30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits
5 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their
10 interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
15 related to SEQ ID NO:51 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
20 formula of a-b, where a is any integer between 1 to 2045 of SEQ ID NO:51, b is an integer of 15 to 2059, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:51, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

25 The translation product of this gene shares sequence homology with protocadherins, which are related to cadherin, and possess cell adhesive ability. Cadherins are glycosylated integral membrane proteins that are involved in cell-cell adhesion.

This gene is expressed primarily in brain (infant, adult frontal lobe, manic
30 depression tissue) and to a lesser extent in epididymus, healing groin wounds, ovary, adipocytes, and fetal tissue (e.g., kidney and retina).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired male and female fertility, developmental disorders, fibrosis, and manic depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 165 as residues: Val-35 to Lys-41, Ser-68 to Gln-73, Glu-88 to Glu-93, Arg-156 to Gly-163, Ala-199 to Gly-206, Asp-216 to Ser-226, Thr-249 to Asn-254, Asp-339 to Pro-345, Ile-370 to Gly-379, Pro-429 to Glu-434, Arg-461 to Pro-466, Ala-475 to Thr-482, Pro-585 to Gly-593, Glu-631 to Gln-639, Pro-674 to Pro-682, Gln-715 to Gly-720, Ser-736 to Arg-742. Polynucleotides encoding said polypeptides are also provided.

BLAST analysis reveals high homology to protocadherin sequences. These sequences are related to cadherin, and possess cell adhesive ability. Such proteins may have regulatory functions in the cell, as well as the cell-cell adhesive properties. Antibodies produced against these sequences are useful for modulating the binding activity of these protocadherins, and can be used therapeutically. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease.

Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive
5 disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,
10 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment
15 of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues
20 of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Moreover, the expression within fetal tissue (e.g., kidney and retina) and other cellular sources marked by proliferating cells indicates this
25 protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including blindness, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions
30 involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:52 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3268 of SEQ ID NO:52, b is an integer of 15 to 3282, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:52, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

Preferred polypeptides of the invention comprise the following amino acid sequence:

IRHEQQGEEDDEHARPLAESLLLAIDLLFCPDFTVQSHRRSTVDSAEDVHSL
 DSCEYIWEAGVGFAHSPQPNYIHDMNRMELLKLLLTFCSEAMYLPPAPESGS
 TNPWVQFFCSTENRHALPLFTSLLNTVCAYDPVGYGIPYNHLLFSDYREPLVE
 EAAQVLIVTLDHDSASSASPTVDGTTTGTAMDDADPPGPENLFVNYLSRIHRE
 5 EDFQFILKGIARLLSNPLLQTYLPNSTKKDPVPPGAASSLLEALRLQQEIPLLRA
 EEQRRPRHPCPHLLPQRCPGRSV (SEQ ID NO: 317). Polynucleotides encoding
 such polypeptides are also provided.

This gene is expressed primarily in brain, breast, breast cancer tissue and to a lesser extent in epididymus, amniotic cells, and embryo tissue.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired CNS function, male sterility, and breast cancer. Similarly, polypeptides and antibodies directed to these
 15 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, male reproductive, cancerous and wounded tissues) or
 20 bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
 25 epitopes shown in SEQ ID NO: 166 as residues: Pro-22 to Pro-31, Ser-38 to His-43, Asp-74 to Leu-79, Asp-113 to Glu-121, Leu-157 to Val-166, Ala-189 to Arg-196, Gln-206 to Arg-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particularly in the cerebellum, indicates polynucleotides and polypeptides corresponding to this gene are useful for the
 30 detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11,

15. and 18. and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. The expression in the breast tissue may indicate its uses in the diagnosis and/or treatment of breast neoplasia and breast cancers, such as fibroadenoma, papillary carcinoma, ductal carcinoma, Paget's Disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma, as well as juvenile hypertrophy and gynecomastia, mastitis and abscess, duct ectasia, fat necrosis and fibrocystic diseases. Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may

show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions
5 involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of
10 potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types
15 of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new
20 insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or
25 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:53 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically
30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1846 of SEQ ID NO:53, b is an integer of 15 to 1860, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:53, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 44

Contact of cells with supernatant expressing the product of this gene increases the permeability of monocytes to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells.

This gene is expressed primarily in CD34 positive cells derived from human cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; defects in hematopoietic stem and progenitor cells; susceptibility to chemotherapy and irradiation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 167 as residues: Ala-38 to Leu-59, Ala-63 to Thr-71, Lys-82 to Leu-91, Glu-97 to Ser-107, Gln-143 to Ala-149, Ile-153 to Leu-158, Ser-169 to Arg-182. Polynucleotides encoding said polypeptides are also provided.

Elevated expression of this gene product in CD34 positive hematopoietic cells indicates that it is expressed by early stem and progenitor cells of the hematopoietic lineages. Therefore, this may represent a soluble factor that is able to control the survival, proliferation, differentiation, or activation of all hematopoietic lineages, including stem and progenitor cells. Thus, it could be quite useful, for example, in ex vivo expansion of stem cell numbers for hematopoietic disorders or for cancer patients. Alternately, it may represent a factor that influences the hematopoietic microenvironment by affecting stromal cells that release other factors required for hematopoietic development. Additionally, the tissue distribution in CD34 positive cells also indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:54 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:54, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:54, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

This gene is expressed primarily in breast and 12-week old human embryos and to a lesser extent in stomach cancer and liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer; stomach cancer; embryonic defects; hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal
15 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and/or treatment of a variety of disorders. Elevated expression
20 of this gene product in stomach cancer indicates it is useful as a marker or therapeutic target for stomach cancer. Alternately, expression in breast tissue is influenced by the presence or absence of breast cancer tissue, and may thus also serve as a diagnostic marker for this cancer as well. Expression in the developing embryo may correlate with the normal development of human embryos, and expression in the liver is
25 involved in the regulation of normal liver function and/or liver regeneration.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:55 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically
30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1079 of SEQ ID NO:55, b is an integer of 15 to 1093, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:55, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in human hypothalamus derived from a patient with schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, schizophrenia; neurological disorders; impaired nervous system function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 169 as residues: Glu-34 to Trp-39. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particularly in the hypothalamus, indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal

cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated
5 expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity,
10 to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:56 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 618 of SEQ ID NO:56, b is an integer of 15 to 632, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:56, and where b is greater than or equal to a + 14.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 47**

The translation product of this gene shares sequence homology with human lecithin-cholesterol acyltransferase (LCAT), which catalyses the transfer of fatty acid from the sn-2 position of lecithin to the free hydroxyl group of cholesterol. Preferred polypeptides of the invention comprise the following amino acid sequence: RLVYN
30 KTSRATQFPDGVDRVPGFGKTFSLFLDPSKSSVGSYFHTMVESLVGWGYT
RGEDVRGAPYDWRRAPNENGPYFLALREMIEEMYQLYGGPVVLVAHSMGN
MYTLYFLQRQPQAWKDKYIRAFVSLGAPWGGVAKTLRVLASGDNNRIPVIG

PLKIREQQRSAVSTSWLLPYNYSPEKVFVQTPTINYTLRDYRKFFQDIGFE
DGWLMRQDTEGLVEATMPPGVQLHCLYGTGVPTPDSFYYESFPDRDPKICFG
DGDGTVNLKSALQCQAWQSRQEHQVLLQELPGSEHIEMLANATTLAYLKRV
LLGP (SEQ ID NO: 318). Polynucleotides encoding such polypeptides are also

5 provided.

This gene is expressed primarily in osteoblasts & dendritic cells and to a lesser extent in muscle and other hematopoietic cell lineages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
10 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; osteoporosis; osteopetrosis; muscle degeneration. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
20 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 170 as residues: Cys-65 to Ser-71. Polynucleotides encoding said polypeptides are also provided.

25 The tissue distribution and homology to lecithin-cholesterol acyltransferase (LCAT) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. For example, atherosclerosis is a pathological condition of mammals characterised by the accumulation of cholesterol in the arteries, which leads to heart disease, strokes, heart
30 attacks and peripheral vascular disease. The enzyme could be used in a novel method of treating atherosclerosis, which involves increasing the level of LCAT activity, which then causes a decrease in the accumulation of cholesterol. The method and the

products can be used for the prophylaxis and treatment of atherosclerosis, and associated heart disease, myocardial infarction, stroke and peripheral vascular disease, as well as individuals suffering from Fish Eye Syndrome (caused by LCAT deficiency) or Classic LCAT Deficiency Syndrome. Alternately, elevated expression of this gene product in osteoblasts and hematopoietic cell lineages indicates that it may play additional roles in bone turnover, regulation of immune system function, and muscular function.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:57 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2673 of SEQ ID NO:57, b is an integer of 15 to 2687, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:57, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

When tested against HELA epithelial cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates epithelial cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in adult brain, infant brain, fibroblasts, embryonic and fetal tissue (e.g., spleen, liver), placenta and to a lesser extent in endocrine organs, cancerous colon and breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, dementia, epilepsy, schizophrenia, and developmental abnormalities.

5 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, endocrine system, and during development, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types
10 (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly,
20 the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive
25 compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,
30 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. In addition, the expression of this gene product in synovium (synovial sarcoma) would suggest a role in the detection and treatment of disorders and

conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). The tissue distribution in endocrine tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Additionally, the expression within fetal tissue, cancerous colon and breast, and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of
5 degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue
10 markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:58 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 605 of SEQ ID NO:58, b is an integer of 15 to 619, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:58; and where b is greater than or equal to a + 14.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 49**

Preferred polypeptides of the invention comprise the following amino acid sequence or a subfragment thereof: MNKEDKVWNDCKGVNKLTNLEEQYIILIFQ
NGLDPPANMVFESIINEIGIKNNISNFFAKIPFEEANGRLVACTRITYEESIKGSC
GQKENKIKTVSFESKIQLRSKQEFQFFDEEEETGENHTIFIGPVEKLIVYPPPPA
30 KGGISVTNEDLHCLNEGEFLNDVIIDFYLYLVLEKLKKEDADRIHIFSSFFYK
RLNQRERRNHETTNLISIQQRHGRVKTWTRHVDIFEKDFIFVPLNEAAHWFL
AVVCFPGLEKPKYEPNPHYHENA VIQKCSTVEDSCISSASEMESCSQNSSAK

PVIKKMLNKKHCIAVIDSNPGQEESDPYKRNICSVKYSVKKINHTASENEEF
NKGESTSQKS (SEQ ID NO: 319). Polynucleotides encoding such polypeptides are also provided.

5 This gene is expressed primarily in fetal tissue, stomach, brain, endometrial cells, and bone and to a lesser extent in prostate, retina, adipocytes, smooth muscle, and tumors of the endometrium, ovaries, and parathyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
10 not limited to, disorders of the endocrine system, ulcers, stomach cancer, epilepsy, schizophrenia, dementia, bone growth, developmental disorders and resorption. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the
15 digestive system and neural systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression
20 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.
25 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia,
30 trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS,

psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation.

5 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Expression of this gene product in stomach tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of stomach cancer, and cancer in general. The expression within embryonic, fetal tissue and other cellular
10 sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere
15 herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain
20 neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating,
25 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in
30 proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in parathyroid tumor indicates polynucleotides and polypeptides corresponding to this

gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. The tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:59 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1364 of SEQ ID NO:59, b is an integer of 15 to 1378, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:59, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

The translation product of this gene shares good protein homology with Xenopus NaDC-2 gene and a rabbit renal sodium/dicarboxylate cotransporter. The translation product of this gene also shares good homology with a rat placental protein which is a sodium-coupled high affinity dicarboxylate transporter. Therefore, it is likely that the translated product encoded by this gene shares similar biological activity.

This gene is expressed primarily in the placenta and colon adenocarcinoma. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental abnormalities as well as failure to thrive anomalies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system and colon, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 173 as residues: Lys-166 to Gly-181. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human placenta and the shared homology of this translation product to a rat placental protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product is produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or

survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product is produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. The tissue distribution in colon tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders involving the colon. Expression of this gene product in colon tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of colon cancer, and cancer in general.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:60 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1112 of SEQ ID NO:60, b is an integer of 15 to 1126, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:60, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in the spinal cord.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, paralysis, neurologic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of

the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken
5 from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in spinal cord indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or
10 prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,
15 Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including
20 disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or
25 survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:61 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
5 formula of a-b, where a is any integer between 1 to 2064 of SEQ ID NO:61, b is an integer of 15 to 2078, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:61, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

10 This gene is expressed primarily in keratinocytes, brain, fetal tissues, pericardium, stomach, and cancerous tissues (e.g., stomach, adrenals, parathyroid, germ cell, colon, breast).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin disorders, neurodegenerative and developmental disorders, heart disease, and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
20 tissues or cells, particularly of the cardiovascular and gastrointestinal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the
25 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.
30 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of

Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive
5 compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,
10 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and
15 "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis
20 fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may
25 predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and
30 inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita,

familial osteoarthritis. Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). The expression within fetal tissue (e.g., spleen and liver) and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

- 10 Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Additionally, the tissue distribution in the pericardium of the heart indicates that the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the
- 15
20
25
30

protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:62 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 748 of SEQ ID NO:62, b is an integer of 15 to 762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:62, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene is expressed primarily in the brain and in cartilage and to a lesser extent in the retina, activated T-cells, pineal gland, the lungs, and in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological diseases, such as epilepsy and dementia, osteoarthritis, retinopathies, hematopoietic diseases, emphysema, and lung cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurologic system, cartilage and musculature, vision, the hematopoietic system, and the pulmonary system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 176 as residues: Arg-34 to Cys-44. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in T-cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such

as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The expression of this gene product in synovium would suggest a role in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). Additionally, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have

applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:63 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1080 of SEQ ID NO:63, b is an integer of 15 to 1094, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:63, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

This gene is expressed primarily in umbilical vein endothelial cells induced by IL-4.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, angiogenesis, inflammatory disorders, hematopoietic disease.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the
5 angiogenic and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy
10 tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in endothelial cells indicates polynucleotides and polypeptides corresponding to this gene are useful in the detection, treatment, and/or prevention of vascular conditions, which include, but are not limited to, microvascular disease, vascular leak syndrome, aneurysm, stroke, atherosclerosis, arteriosclerosis, or
15 embolism. For example, this gene product may represent a soluble factor produced by smooth muscle that regulates the innervation of organs or regulates the survival of neighboring neurons. Likewise, it is involved in controlling the digestive process, and such actions as peristalsis. Similarly, it is involved in controlling the vasculature in areas where smooth muscle surrounds the endothelium of blood vessels. Furthermore,
20 the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. The secreted protein can also be
25 used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere
30 herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human

immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility);
5 chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the
10 a corresponding nucleic acid in gene therapy procedures.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:64 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically
15 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1347 of SEQ ID NO:64, b is an integer of 15 to 1361, where both a and b correspond to the positions of nucleotide
20 residues shown in SEQ ID NO:64, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene is expressed primarily in both normal and cancerous pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diabetes, gastrointestinal disorders, and pancreatic cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
30 a number of disorders of the above tissues or cells, particularly of the digestive and blood systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution in pancreas indicates that the protein products of this gene are useful as a therapeutic and/or diagnostic agent for pancreatic disorders and disorders of the endocrine and exocrine system, including but not limited to diabetes, blood disorders, pancreatic cancer, gastrointestinal diseases, hormonal imbalance, autoimmune disorders, cystic fibrosis, pancreatitis, and gallstones. Furthermore, the
10 protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:65 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
20 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 933 of SEQ ID NO:65, b is an integer of 15 to 947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:65, and where b is greater than or equal to a + 14.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

The translation product of this gene shares sequence homology with oxidoreductase. Preferred polypeptides of the invention comprise the following amino acid sequence: MSPLSAARAALRVYAVGAAVILAQLLRRCRGGFLEPVXPPRP
30 DRVAIVTGGTDGIGYSTANIWRDLGMHVIIAGNNSKAKQVVSKIKEETLND
KVEFLYCDLASMTSIRQFVQKFKMKKIPLHVLINNAGVMMVPQRKTRDGFE
HFGLNYLGHFLLTNLLDNLKESGSPGHSARVVTVSSATHYVAELNMDDLQS

SACYSPHAAYAQSKLALVLFTYHLQRLLAAEGSHVTANVVDPGVVNTDXYK
HVFWATRLAKKLLGWLLFKTPDEGAWTSIYAAVTPELEGVGGRYLYNEKET
KSLHVTYNQKLQQQLWSKSCMTGVLDVTL (SEQ ID NO: 320). The mature
form of this protein begins at residue 32. Thus, polypeptides comprising residues 2-
5 330 and 32-330 of the sequence shown above are also provided. Polynucleotides
encoding such polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following
amino acid sequence: MSPLSAARAALRVYAVGAAVILAQLLRRCRGGFLEP
VXPPRPDRVAIVTGGTDGIG YSTANIWRDLACMLS (SEQ ID NO: 321).

10 Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in breast cancer cells, osteoclastoma, wilm's
tumor, thymus stromal cells, and T cell helper I.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, cancer, e.g., breast cancer, osteoclastoma, and wilm's tumor. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
a number of disorders of the above tissues or cells, particularly of the immune system,
20 expression of this gene at significantly higher or lower levels is routinely detected in
certain tissues or cell types (e.g., reproductive, kidney, immune, hematopoietic, and
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, breast milk,
urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
25 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution in breast cancer tissue, combined with the homology to
oxidoreductase indicates that polynucleotides and polypeptides corresponding to this
gene are useful for diagnosis and treatment of cancer, particularly, breast cancer,
30 osteoclastoma, and wilm's tumor. This protein may play a role in the regulation of
cellular division, and may show utility in the diagnosis, treatment, and/or prevention
of developmental diseases and disorders, including cancer, and other proliferative

conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

5 Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have
10 applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue
15 differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the
20 protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:66 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
30 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1362 of SEQ ID NO:66, b is an

integer of 15 to 1376, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:66, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

5 This gene is expressed primarily in monocytes, T cell helper II and B cell lymphoma.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly B-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 180 as residues: Asp-30 to Val-40. Polynucleotides encoding said polypeptides are also provided.

 The tissue distribution in monocytes, T cell helper, and B cell lymphoma cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell lymphoma. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for

5 immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

10 disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in

15 the expansion of stem cells and committed progenitors of various blood lineages; and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

20 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:67 and may have been publicly available prior to conception of

25 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2420 of SEQ ID NO:67, b is an

30 integer of 15 to 2434, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:67, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene is expressed primarily in human lung cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, pulmonary diseases and/or disorders, particularly cancers of the lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary lavage, pulmonary surfactant, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 181 as residues: Phe-39 to Asp-45. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in lung cancer tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as ARDS, cystic fibrosis, and cancer, particularly lung cancer. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:68 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1072 of SEQ ID NO:68, b is an integer of 15 to 1086, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:68, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in larynx carcinoma and early stage human lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, gastrointestinal, and pulmonary diseases and/or disorders, particularly larynx carcinoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, gastrointestinal, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, pulmonary lavage, sputum, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 182 as residues: His-42 to Lys-49. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in larynx carcinoma and early stage human lung indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating immune system disorders such as cancer, particularly larynx carcinoma. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions. in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of

5 degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue

10 markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly

15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:69 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1248 of SEQ ID NO:69, b is an integer of 15 to 1262, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:69, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 60

Preferred polypeptides of the invention comprise the following amino acid sequence: MEVTTEDTSRTDVSEPATSGGAADGVTSIAPTAVASSTTAASITTA

ASSMTVASSAPTTAASSTTVASIAPTTTASSMTAASSTPMTLALPAPTSTXTGR

TPSTTATGHPSLSTALAQVPKSSALPRTATLATLATRAQTVATTANTSSPMST

30 RPSPSKHMPSDTAASPVPPMXPQAQGPISQVSVDQPVVNTTXKSTXMPSNTT

XEPLTQAVVDKTLVVLLLVGVTLFITVLVLFALQAYESYKKKDYTQVDYLI

NGMYADSEM (SEQ ID NO: 322). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARCPGLPGLRCRPRPRAGPQAPSYCPRATRPPG ACCARMRLLEWRVYLRLTCATKDGMARECPTTWLSPPAKPDFAQRHSVK PTALQGGRWSRLGASP (SEQ ID NO: 323). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in adipocytes, osteoblasts, cerebellum, hypothalamus and Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic, skeletal, neural, and immune diseases and/or disorders, particularly Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, skeletal, neural, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 183 as residues: Pro-33 to Gln-40, Gly-51 to Arg-56. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in Hodgkin's lymphoma cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as cancer, particularly Hodgkin's lymphoma. The secreted protein can also be used to determine biological activity, to

raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:70 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1628 of SEQ ID NO:70, b is an integer of 15 to 1642, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:70, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

The translation product of this gene shares sequence homology with polypeptide in the cystatin family. Cystatin polypeptides are cysteine protease inhibitors. For an analysis of the composition of several members of the cystatin family see Gene (1987) 61(3):329-338, incorporated herein by reference. The cystatin activity of polypeptides encoded by this gene is measured by several assays known in the art including assays described in coowned, copending US Patent Application Serial No. 08/744,138, incorporated herein by reference. Preferred polypeptides of the invention comprise the following amino acid sequence: LPATVEFAVHTFNQQSKD
10 YYAYRLGHILNSWKEQVESKTVFSMELLGRTRCGKFEDDIDNCHFQESTEL
NNTFTCFFTISTRPWMTQFSLLNKC (SEQ ID NO: 324). Fragments of such polypeptides having cystatin activity (cysteine protease inhibitory activity are particularly preferred). Polynucleotides encoding such polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of
15 the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LLWARGLGRAKSAVPTVST MLGLPWKGGLS
WALLLLLLGSQILLIYAWHFHEQRDCDEHNVMARYLPATVEFAVHTFNQQS
KDYYAYRLGHILNSWKEQVESKTVFSMELLGRTRCGKFEDDIDNCHFQE
20 STELNNTFTCFFTISTRPWMTQFSLLNKTCLGFH (SEQ ID NO: 325).
Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in testes and epididymus. For a review of a cystatin showing testes- specific expression see Mol. Endocrinol. (1992 Oct.)
6(10):1653-1664, incorporated herein by reference.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled
30 proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes

in rheumatoid arthritis, purulent bronchiectasis and periodontitis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune,

- 5 expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, seminal fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in
10 healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 184 as residues: Phe-31 to Asp-38, Asn-59 to Tyr-65, Ser-76 to Glu-82, Thr-96 to Cys-108, Gln-111 to Asn-118. Polynucleotides encoding said polypeptides are also provided.

- 15 The tissue distribution in testes and epididymus, combined with the homology to cystatins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or
20 impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene
25 product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues
30 rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Cysteine proteinase inhibitors of the cystatin superfamily are ubiquitous in the body and are generally tight-binding inhibitors of papain-like cysteine proteinases, such as cathepsins B, H, L, S, and K. They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes in rheumatoid arthritis, purulent bronchiectasis and periodontitis, which indicates that a tight cystatin regulation is a necessity in the normal state. A deficiency state in which the levels of the intracellular cystatin, cystatin B, are lowered due to mutations has recently been shown to segregate with a form of progressive myoclonus epilepsy, which points to additional specialized functions of cystatins. Moreover, results showing that chicken cystatin inhibits polio virus replication, human cystatin C inhibits corona- and herpes simplex virus

replication, and human cystatin A inhibits rhabdovirus-induced apoptosis in cell cultures indicates that cystatins play additional roles in the human defense system. The cystatins constitute a superfamily of evolutionarily related proteins, all composed of at least one 100-120 residue domain with conserved sequence motifs.

5 The previously well characterized single-domain human members of this superfamily could be grouped in two protein families. The Family 1 members, cystatins (or stefins) A and B, contain approximately 100 amino acid residues, lack disulfide bridges, and are not synthesized as preproteins with signal peptides. The Family 2 cystatins (cystatins C, D, S, SN, and SA) are secreted proteins of approx.
10 120 amino acid residues (Mr 13,000-14,000) and have two characteristic intrachain disulfide bonds. Recently, we identified an additional human cystatin superfamily member by EST1 sequencing in epithelial cell derived cDNA libraries which we named cystatin E. The same cystatin was independently discovered by differential display experiments as a mRNA species down-regulated in breast tumor tissue, but
15 present in the surrounding epithelium and reported under the name cystatin M. Cystatin E/M is an atypical, secreted low-Mr cystatin in that it is a glycoprotein and just shows 30-35% sequence identity in alignments with the human Family 2 cystatins, which shows that additional cystatin families are yet to be identified. The cystatin E/M gene has been localized to chromosome 2, whereas all human Family 2
20 cystatin genes are clustered on the short arm of chromosome 20, which further stresses that cystatin E/M is just distantly related to the other secreted human low-Mr cystatins. It is believed therefore, that polypeptides encoded by this gene are useful in diagnosing and treating disease consistent with the aforementioned conditions in which cystatins are implicated.

25 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:71 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
30 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 907 of SEQ ID NO:71, b is an

integer of 15 to 921, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:71, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

- 5 The translation product of this gene shares sequence homology with Neutrophil Gelatinase-Associated Lipocalin which is thought to be important in immune regulation (See Genbank and Geneseq Accession Nos. emb1CAA58127.1, and US5627034, respectively; all references and information available through these accessions are hereby incorporated herein by reference; for example, Biochem. Biophys. Res. Commun. 202 (3), 1468-1475 (1994), and FEBS Lett. 314 (3), 386-388 (1992)).

- In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LEQKLELHRGGGRSRTSGSPGLQEFGTREERGE
 15 GEQRTGREFSGNGGRAVEAARMRLLCGLWLWLSLLKVLQAQTPTPLPLPP
 PMQSFQGNQFQGEWFLGLAGNSFRPEHRALLNAFTATFELSDDGRFEVWN
 AMTRGQHCDTWSYVLIPAAQPGQFTVDHGVGRSWLLPPGTLDQFICLGRAQ
 GLSDDNIVFPDVTGXALDL XSLPWVAAPA (SEQ ID NO: 326). Polynucleotides
 20 encoding these polypeptides are also provided.

This gene is expressed primarily in epididymus and osteoclastoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
 25 not limited to, reproductive and skeletal diseases and/or disorders, particularly cancers such as osteoclastoma testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene
 30 at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, skeletal, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, urine, synovial fluid and spinal fluid)

or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
5 epitopes shown in SEQ ID NO: 185 as residues: Met-82 to Thr-90. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididymus and homology to neutrophil gelatinase-associated lipocalin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of skin diseases and immune system
10 disorders such as cancer, particularly osteoclastoma. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity"
15 sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy);
20 regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction
25 etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or
30 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:72 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 892 of SEQ ID NO:72, b is an integer of 15 to 906, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:72, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 63

The translation product of this gene was shown to have homology to colipase which plays an essential role in the intestinal fat digestion by anchoring lipase on lipid/water interfaces in the presence of bile salts (See Genbank Accession No. gb|AAA03513.1; all references and information available through this accession are
15 hereby incorporated by reference herein).

This gene is expressed primarily in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
20 not limited to, reproductive diseases and/or disorders, particularly epididymus-related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or
25 lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, bile, chyme, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from
30 an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 186 as residues: Ile-40 to Cys-49, Arg-52 to Cys-57,

Ser-94 to Trp-99, Gly-105 to Gly-111. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system diseases and disorders of the epididymus. Polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:73 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 666 of SEQ ID NO:73, b is an integer of 15 to 680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:73, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 64

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: MCVCKERKRGREKEGGVTPTMTSNFPFCTLILGI
5 AQAQACPGCPGDWPGLGSGVGEGLHHIRTCRTPIPCSPAPAAACLGSGH
ARLPCVLRLWPVPANLSSPFRLEALHCSFWSSPLLPAHLAFFGFRDLLTDFL
LAACLLTFQKTPLELPMVAVVHLLVATPCYQMLDNLPLPSAAAN WC (SEQ ID
10 NO: 327). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in melanocytes and placenta and to a lesser extent in bone marrow and many cells of the immune system, including B-cells, dendritic cells, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin cancer and disorders of the reproductive and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
15 type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., reproductive tissue, hematopoietic tissue, melanocytes and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic
20 fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in melanocytes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the skin, the reproductive system, and the immune system,
30 particularly cancers. Representative uses are described in the "Biological Activity",

"Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:74 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1619 of SEQ ID NO:74, b is an integer of 15 to 1633, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:74, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Preferred polypeptides of the invention comprise the following amino acid sequence: YLWGRPRLRM RAGTSPSAPWG EKREKLGHKLPVALQGYHPWIL
LECTVFWARVVLACFSLYLIRGPN CINTRQPEPTYQKACNLDCSSDFGQER
APAWELLGPESEQRLREYTAQGLQSLASSHRWRQFKTEGKMRGGASPLPWL I
10 CFW LCSYKGS DNSLKP VVPGPTLCPQSLVSPSVHPSTRSASLGRHRAEAA
(SEQ ID NO: 328). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the
15 following amino acid sequence: MPGILAGIPVKDLCLSL LQGFRLLLLCVC PGWL
SGWMGGQKGSPRIVDIG (SEQ ID NO: 329). Polynucleotides encoding these polypeptides are also provided. This gene maps to chromosome 15, accordingly, polynucleotides of the invention is used in linkage analysis as a marker for chromosome 15.

20 This gene is expressed primarily in brain and breast and to a lesser extent in the liver, pancreas, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
25 not limited to, disorders affecting the brain and CNS, the reproductive system, or the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, the reproductive system, and the immune
30 system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., brain and other tissue of the nervous system, mammary tissue, endocrine tissue, hepatic tissue, reproductive tissue, cells

and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 188 as residues: Met-37 to Ser-43. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the central nervous system, the reproductive system, and the immune system, including cancers. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:75 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1008 of SEQ ID NO:75, b is an integer of 15 to 1022, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:75, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 66

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AKGEERKEAFSLKMVQLSSEPISFGLMYLYLGV
15 FFHLYPGALSITTLGKHSHPFFTAEQNSTVWMEHTLFHQSPVASHLVCFQSF
AFSE (SEQ ID NO: 330). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the brain and the immune system, in particular T-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting the brain, such as Alzheimer's or disorders affecting the immune system, such as AIDS. Similarly, polypeptides and antibodies directed to
25 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and CNS and the immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, cells and
30 tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene.

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the brain and CNS or disorders affecting the immune system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:76 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1170 of SEQ ID NO:76, b is an

integer of 15 to 1184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:76, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

5 The translation product of this gene shares sequence homology with penaeidin-2 which is thought to be a members of a new family of antimicrobial peptides from the hemolymph of shrimps *Penaeus vannamei*. The molecules display antimicrobial activity against fungi and bacteria with a predominant activity against Gram-positive bacteria.

10 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPAHPASPPLMTLSLQLAELVHFVCAFAQSQWT GVYPMPPLKPTPLCFA CVPCR (SEQ ID NO: 331). Polynucleotides
15 encoding these polypeptides are also provided.

 This gene is expressed primarily in spleen.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
20 not limited to, immune and hematopoietic diseases and/or disorders, particularly disorders affecting the spleen, including bacterial and fungal infections. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic
25 and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, and cells and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or
30 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in spleen and homology to the penaeidin family of antibiotics indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the spleen, especially fungal and bacterial infections. Representative uses are described in the "Immune
5 Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or
10 other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such
15 as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic
20 lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and
25 in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or
30 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:77 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 298 of SEQ ID NO:77, b is an integer of 15 to 312, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:77, and where b is greater than or equal to a + 14.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal
15 transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating immune and hematopoietic cells and tissue cell types. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane
20 potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

Moreover, when tested in TF-1 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: c-fos, c-jun, egr-1, b561, bcl-2, CD40, cyclin D2, GADPH, ICER, MAD3, p21, STAT3,
25 ID3, and STAT-1. When tested in U937 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: egr2, MKP1, ATF3, B562, cyclin D, cyclin D2, GATA3, MAD3, p21, TGF, DHFR, and JAK3. Based upon these results, it is anticipated that polynucleotides and polypeptides corresponding to this gene are useful as agonists or antagonists of the
30 above referenced genes. Such activity is useful in therapeutic and/or diagnostic applications as referenced and more specifically discussed elsewhere herein.

In specific embodiments, polypeptides of the invention comprise the sequence: MLLEVYGDSISVTVAIPL (SEQ ID NO: 332), MHSPCQSKAADGLGKSETE (SEQ ID NO: 333), and/or MLKSLGLSTN (SEQ ID NO: 334). Polynucleotides encoding these polypeptides are also provided.

5 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AQRLAEECFYMLLEVYGDSISVTVAIPLMHSP
CQSKAADGLGKSETEMLKSLGLSTNMSPFHLLGLKVFLTWALTLAQICLY
10 FFEVQPLGLLALNFFCTATAGLKELCMHPPSLAFTPEFHTSLSPAIPSFCGTS
VLSNSHTIPLSLYLFPFSKSRMPDTLHLLVHSLPLVHSQVLPVKDVTIEWPLC
QRCLGSTCH Q (SEQ ID NO: 335). Polynucleotides encoding these polypeptides are also provided.

The polypeptide of this gene has been determined to have a transmembrane
15 domain at about amino acid position 11 - 27 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 28 to 143 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

20 This gene is expressed primarily in neutrophils and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders and/or diseases affecting the immune system. Similarly,
25 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, cells and tissue
30 of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 191 as residues: Pro-97 to Asp-104. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in neutrophils and T-cells, combined with the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation, survival, differentiation, and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their

interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:78 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1356 of SEQ ID NO:78, b is an integer of 15 to 1370, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:78, and where b is greater than or equal to a + 14.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 69**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEVQVSLFQMFCFSSIFCSH
20 EHTHLPGTFWLFLFLILPPSCPCFLPFS LAIETVRWPCWHHPTSFELCY
PGTSIYYASRGGPXPNSEX (SEQ ID NO: 336). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases and/or disorders affecting the immune system, and neutrophils in particular. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the
30 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., blood cells,

and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system and neutrophils in particular. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as:

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:79 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 354 of SEQ ID NO:79, b is an integer of 15 to 368, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:79, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: XNXKSPLTIGNKSWSTAVAALELVDPGCR NSARDSPELVHLGKGRPRKLMTYLFCSSISLLLLKVHSSGHQDIRKAKSKVP RLLIIQCPQQRE (SEQ ID NO: 337). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting smooth muscle tissue, particularly vascular conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of smooth muscle tissue expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 193 as residues: Ser-18 to Val-31. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution primarily in smooth muscle indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting smooth muscle tissue. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:80 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1074 of SEQ ID NO:80, b is an integer of 15 to 1088, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:80, and where b is greater than or equal to a + 14.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by

the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPEENLSPSTPSQMPTIWVKLCLLQVCHGLFP LLKHWSQPMPLCVTLAPVSYWL (SEQ ID NO: 338). Polynucleotides encoding these polypeptides are also provided.

5 This gene is expressed primarily in fetal heart, smooth muscle, and frontal cortex.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
10 not limited to, muscular, vascular, or neural diseases and/or disorders, particularly defects or injury to cardiac muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at
15 significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscular, vascular, neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy
20 tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution in fetal heart indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating defects to the heart either due to injury or congenital defects. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and
25 conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory
30 conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection,

treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:81 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1848 of SEQ ID NO:81, b is an integer of 15 to 1862, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:81, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translation product of this gene shares sequence homology with adipose complement related protein which is thought to be important in regulating energy metabolism, insulin levels and fat stores. Moreover, the protein product of this gene has also been shown to have homology to the complement subcomponent C1q A-chain precursor and HP-25 protein (See Genbank and Geneseq Accession Nos. emb|CAA41664.1, dbj|BAA02352.1, and W98013; all references and information

available through this accession are hereby incorporated by reference herein). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with complement proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PRVRKEPEAMQWLRVRESPGEATGHRVTMG TAALGPVWAALLLFLLMCEIPMVELTFDRAVASDCQRCCDSEDPLDPAHVSS ASSSGRPHALPEIRPYINITILKDGKDGPMGLPGYMGREGPGQGPQGSGK
10 GDKGEMGSPGAPCQKRFFAFSVGRKTALHSGEDFQTLLFERVFVNLDGC
FDMATGQFAAPLRGIYFFSLNVHSWNYKETVHIMHNQKEAVILYAQPS
ERSIMQSQSVMLDLAYGDRVWVRLFKRQRENAIYSNDFDTYITFSGHLIKA
EDD (SEQ ID NO: 339). Polynucleotides encoding these polypeptides are also provided.

15 This gene is expressed primarily in placenta and, fetal kidney, and umbilical vein and to a lesser extent in fetal heart, fetal liver/spleen, microvascular endothelial cells and cancers of the lung and pharynx.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, vascular, renal, and reproductive diseases and/or disorders, particularly cancers of the lung and pharynx. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
25 tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, renal, reproductive, immune, hematopoietic, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having
30 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 195 as residues: Asp-36 to Asp-48, Ser-57 to His-62, Lys-77 to Gly-84, Met-92 to Gly-114, Gln-203 to Ile-209, Lys-231 to Tyr-239. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in pharynx or lung, combined with the homology to adipose complement related proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancers of the pharynx or lung by modifying the metabolic balance in such tissues. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, 10 vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product 15 may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional 20 supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:82 and may have been publicly available prior to conception of 25 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1604 of SEQ ID NO:82, b is an 30 integer of 15 to 1618, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:82, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The translation product of this gene shares sequence homology with a hypothetical 54.7 kD protein (F37A4.1) from *Caenorhabditis elegans* (SwissProt locus YPT1_CAEEL, accession P41879). The protein product of this gene also has
5 homology to the human NG26 which is thought to contain a human major histocompatibility complex class III and is involved in T-cell maturation (See Genbank Accession No. gblAAD18079.11 (AF129756); all references and information available through this accession are hereby incorporated by reference herein; for example, J. Neurochem. 69 (6), 2516-2528 (1997)). Based on the sequence similarity,
10 the translation product of this gene is expected to share at least some biological activities with nitric oxide synthase proteins.

Preferred polypeptides of the invention comprise the following amino acid sequence: MLYPGSVYLLQKALMPVLLQGQARLVEECNGRRAKLLACDGNE
IDTMFVDRRGTAEPQGQKL VICCEGNAGFYEVGCVSTPLEAGYSVLGWNHP
15 GFAGSTGVPPQNEANAMDVVVQFAIHR LGFQPQDIIYAWSIGGFTATWAA
MSYPDVSAMILDASFDDL VPLALKVMPDSWRGLVTRTVRQHLN LNNAEQLC
RYQGPVLLIRRTKDEIITTTVPEDIMSNRGN D LLLKLLQHRYPRVMAEEGLRV
VRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGPDFPWSVGEDMS
ADGRRQLALFLARKHLHNFEATHCTPLPAQNFQMPWHL (SEQ ID NO: 340).
20 Polynucleotides encoding such polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VCPKWCRFLTMLGHCCYFWQVWPASEALAA
25 GPTPSTGSSSPSWKQHIGTSLQKTRGSLPTTTLTSGAGQSTSTGKNPAAGR
SLEGALPAGVWPCFAQSPCTGGQQT P SSTGLRSCLVRSPATWW RTP (SEQ ID NO: 341). Polynucleotides encoding these polypeptides are also provided.

Preferred polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEIYREXDSERAPASVPETPTAVTAPHSSSWDTYYQ
30 PRALEKHADSILALASVFW SISYSSPFAFFYL YRKG YLSLSKVVPFSHYAG
TLLLLLAGVACXRIGRW TNPQYRQFITILEATHRNQSS ENKRQLANYNFD
FRSWPVDFHWEPPSSRKESRGGPSRRGVALLRPEPLHRGTADTLLNRVKKL

PCQITSYLV AHTLGRRLYPGSVYLLQKALMPVLLQGQARLVEECNGRRRAK
LLACDGNEIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTPLEAGYS
VLGWNHPGFAGSTGVPFPQNEANAMDVVVQFAIHR LGFQPQDIIYAWSI
GGFTATWAAMSYPDVSAMILDASFDDLVP LALKVMPDSWRGLVTRTVRQ
5 HLN LNNAEQLCRYQGPVLLIRRTKDEIITTTVPEDIMSNRGNDLLLKLLQHR Y
PRVMAEEGLRVVRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGP
DFPWSVGEDMSADGRRQLALFLARKHLHNFEATHCT PLPAQNFQMPWHL

(SEQ ID NO: 342). Polynucleotides encoding these polypeptides are also provided. A
preferred polypeptide variant of the invention comprises the following amino acid
10 sequence: HERAXGPSRGHGELLSCVLGPRLYKIYRERDSE RAPASVPETPTA
VTAPHSSSWDTYYQP RALEKHADSILALASVFWSISYYSSPFAFFYL YRKG Y
LSLSKVVPFSHYAGTLLLLLAGV ACSEALAAGPTPSTGSSSPSWKQHIGTSLQ
KTRGSLPTTTLTSGAGQSTSTGKNPAAGRSLEGALPAGVWPCFAQSPCTGG
QQTPSSTGL RSCLVRSPATWW RTP (SEQ ID NO: 343). Polynucleotides
15 encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome
6. Accordingly, polynucleotides related to this invention are useful as a marker in
linkage analysis for chromosome 6.

This gene is expressed primarily in cerebellum, pituitary, fetal liver, and
20 primary dendritic cells and to a lesser extent in a wide range of tissues and
developmental stages (i.e. fetal and adult tissue, etc.).

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
25 not limited to, neural, developmental, and immune diseases and/or disorders,
particularly those involving self recognition and T- and B-cell maturation, and cancer.
Similarly, polypeptides and antibodies directed to these polypeptides are useful in
providing immunological probes for differential identification of the tissue(s) or cell
type(s). For a number of disorders of the above tissues or cells, particularly of the
30 neural or hormonal system, expression of this gene at significantly higher or lower
levels is routinely detected in certain tissues or cell types (e.g., neural, developmental,
immune, hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., serum,

plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 196 as residues: Thr-23 to Lys-34, Leu-41 to Ser-47, Ala-57 to Ala-68, Pro-89 to Gly-101, Pro-110 to Pro-117. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in developmental and immune cells, combined with the
10 homology to the human major histocompatibility complex class III region, indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and other proliferative disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the
15 expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

20 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory
bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,
25 such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that
30 influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and

in the differentiation and/or proliferation of various cell types. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to
5 isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
10 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:83 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2020 of SEQ ID NO:83, b is an integer of 15 to 2034, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:83, and where b is greater than or equal to a + 14.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 74

The translation product of this gene shares sequence homology with the br-1 protein from the snail nervous system (EMBL HPBR1GENE) which codes for nitric oxide synthetase and which is thought to be important in mediating a variety of cellular responses, including vasodilation. Preferred polypeptides of the invention
25 comprise the following amino acid sequence: MFKRHQRLKKDSTQAEEDLSEQ
EQNQLNVLKKHGYVVGRVGRFTLYSEEQKDNIPFEFDADSLAFDMENDPVM
GTHKSTKQVELTAQDVKDAHWFYDTPGITKENCILNLLTEKEVNIVLPTQSIV
PRTFVLKPGMVLFLGAIGRIDFLQGNQSAWFTVVASNILPVHITSLDRADALY
QKHAGHTLLQIPMGGERMAGFPPLVAEDIMLKEGLGASEAVADIKFSSAG
30 WVSVTPNFKDRLHLRGYTPEGTVLTVRPPLLPYIVNIKGQRIKKSVA YKTKKP
PSLMYNVRKKKGKINV (SEQ ID NO: 344). Polynucleotides encoding such polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MLPARLPFRLLSLFLRGSAPTAARHGLREPLLERRCAA ASSFQHSSSLGRELPYDPVDTEGFGEGGDMQERFLFPEYILDPEPQPTREKQL QELQQQQEEEEERQRQRREERRQQNLRARSREHPVVGHDPALPPSGVNCSS
5 GCGAXLHCQDAGVPGYLPREKFLRTAEADGGLARTVCQRCWLLSHHRRALR LQVSREQYLELVSAALRXPGPSLVLYMVDLLDLPDALLPDLPALVGPQQLIV LGNKVDLLPQDAPGYRQRLRERLWEDCARAGLLLAPGTKGHSAPSRTSHR TGRIRIRRTGPAQWSGTCTG (SEQ ID NO: 345). Polynucleotides encoding these polypeptides are also provided.

10 When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid cells through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is
15 a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in early stage human brain, smooth muscle,
20 and endometrial tumor and to a lesser extent in a variety of tissues representing many organs and developmental states.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
25 not limited to, cardiovascular, vascular, and neural diseases and/or disorders, particularly congestive heart disease and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and
30 neural systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, vascular, neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 197 as residues: Phe-42 to Leu-48, Pro-53 to Asp-58, Pro-81 to Glu-123, Asp-256 to Trp-269, Gly-282 to Ser-306, Arg-333 to Gly-339, Arg-403 to Gln-425, Ser-446 to Asn-452, His-475 to Gln-480, Gly-592 to Met-597, Pro-635 to His-642, Lys-667 to Lys-672, Lys-678 to Ser-684. Polynucleotides
10 encoding said polypeptides are also provided.

The tissue distribution in smooth muscle and vascular tissues, combined with the homology to nitric oxide synthetase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of congestive heart failure and neurological degenerative disorders. polynucleotides and
15 polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection,
20 treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder,
25 learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,
30 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to

microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis.

Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents
5 that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
10 related to SEQ ID NO:84 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
15 formula of a-b, where a is any integer between 1 to 2226 of SEQ ID NO:84, b is an integer of 15 to 2240, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:84, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

20 The translation product of this gene shares sequence homology with the human KE04p, in addition to an unidentified *C.elegans* gene.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 9 - 25 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 to 8
25 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by
30 the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PSFRRERVETGGGGPVTHGTEGPFLPLPGGTRM NMTQARVLVAADVGLVAVLLYASIHKIEEGHLAVYYRGGALLTSPSGPGYH

IMLPFITFRSVQTTLQTDEVKNVPCGTSGGVMIYIDRIEVVNMLAPYAVFDIV
RNYTADYDKTLIFNKIHHELNQFCSAHTLQEVYIELFDQIDENLKQALQKDL
NLMAPGLTIQAVRVTKPKIPEAIRRFELMEAEKTKLLIAAQKQKVVEKEA
ETERKKAVIEAEKIAQVAKIRFQKQVMEKETEKRISEIEDAAFLAREKAKA
5 DAEYYAAHKYATSNKHKLTPYELELKKYQAIASNSKIYFGSNIPNMFVDSSC
ALKYSD IRTGRESSLPSKEALEPSGENVIQNKESTG (SEQ ID NO: 346).

Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome
10. Accordingly, polynucleotides related to this invention are useful as a marker in
10 linkage analysis for chromosome 10.

This gene is expressed primarily in fetal tissue, including 8 week whole
embryo, fetal liver spleen, nine week old early stage human, fetal heart, fetal liver,
fetal lung, and placenta and to a lesser extent in a variety of cancers, and other normal
tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, cancer and diseases of fetal development. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological
20 probes for differential identification of the tissue(s) or cell type(s). For a number of
disorders of the above tissues or cells, particularly of the fetal tissues, especially the
liver, expression of this gene at significantly higher or lower levels is routinely
detected in certain tissues or cell types (e.g., developmental, hepatic, immune,
hematopoietic, pulmonary, cardiovascular, and cancerous and wounded tissues) or
25 bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another
tissue or cell sample taken from an individual having such a disorder, relative to the
standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
30 epitopes shown in SEQ ID NO: 198 as residues: Leu-68 to Lys-74, Tyr-109 to Lys-
115, Gln-200 to Val-205, Lys-207 to Lys-214, Glu-237 to Ile-244, Ala-271 to Thr-

279, Ser-317 to Ser-329, Gln-342 to Gly-348. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of this gene (primarily fetal tissue and cancerous tissue, both of which are undergoing rapid growth) indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and disorders of fetal development. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:85 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1474 of SEQ ID NO:85, b is an integer of 15 to 1488, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:85, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

When tested against U937 and Jurkat cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid and T-cells, and to a lesser extent in other immune cells and tissue cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WSTGNASWEKKDNFILSADFEMMGLGNRR
SMKSPPLVLAALVACIIVLGFNYWIIASSRSVDLQTRIMELEGRVRRRAAERG
AVELKKNEFQGELEKQREQLDKIQSSHNFQLESVNKLYQDEKAVLVNNITTG
ERLIRVLQDQLKTLQRNYGRLQQDVLQFQKNQTNLERKFSYDLSQCINQMKE
VKEQCEERIEEVTKKGNEAVASRDLSENNDRQQQLQALSEPQPRLQAAGL

PHTEVPQGKGNVLGNSKSQTPAPSSEVVLD SKRQVEKEETNEIQVVNEE
PQRDRLPQEPGREQVVEDRPVGGRGFGGAGELGQTPQVQAALXVSQENPE
MEGPERDQLVIPDGQEEEQEAAGEGRNQKLRGEDDYNMDENEAESETDKQ
AALAGNDRNIDVFNVE DQKRDTINLLDQREKRNHTL (SEQ ID NO: 347).

- 5 Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 9. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 9.

- 10 This gene is expressed primarily in human endometrial tumor and other tumors and to a lesser extent in a variety of other healthy adult and fetal tissues

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental diseases and/or disorders, particularly cancer and other
15 proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrial tissue, cervix and uterus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues
20 or cell types (e.g., developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 199 as residues: Asn-6 to Lys-12, Leu-65 to Phe-70, Glu-73 to His-88, Gln-123 to Gln-135, Gln-142 to Leu-156, Arg-173 to Gly-181, Asp-189 to Gln-199, Ser-204 to Arg-209, Glu-219 to Gly-225, Gly-229 to Pro-238, Ser-246 to Asn-256, Glu-263 to Arg-276. Polynucleotides encoding said polypeptides
30 are also provided.

The tissue distribution in endometrial tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

endometrial, cervical and uterine cancer. Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:86 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3160 of SEQ ID NO:86, b is an integer of 15 to 3174, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:86, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

The translation product of this gene shares sequence homology with protein disulfide isomerase from *Acanthamoeba castellanii* (See Genbank Locus ACADISPROA accession L28174, genpep locus 456013) which is thought to be important in converting proteins into their native conformations. The protein product of this gene was also shown to have homology to a phospholipase C homologue

derived from a mast cell cDNA library (See Geneseq Accession No. R99411). All references and information available through these accessions are hereby incorporated by reference herein - for example, Gene 150 (1), 175-179 (1994).

Included in this invention as preferred domains are endoplasmic reticulum targeting sequence domain and the thioredoxin family active site domain, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Proteins that permanently reside in the lumen of the endoplasmic reticulum (ER) seem to be distinguished from newly synthesized secretory proteins by the presence of the C-terminal sequence Lys-Asp-Glu-Leu (KDEL) [1,2]. While KDEL is the preferred signal in many species, variants of that signal are used by different species. This situation is described in the following table.

Signal	Species-----
KDEL	Vertebrates, Drosophila, Caenorhabditis elegans, plants
HDEL	Saccharomyces cerevisiae, Kluyveromyces lactis, plants
DDEL	Kluyveromyces lactis
ADEL	Schizosaccharomyces pombe (fission yeast)
SDEL	Plasmodium falciparum

The signal is usually very strictly conserved in major ER proteins but some minor ER proteins have divergent sequences (probably because efficient retention of these proteins is not crucial to the cell). Proteins bearing the KDEL-type signal are not simply held in the ER, but are selectively retrieved from a post-ER compartment by a receptor and returned to their normal location. The consensus pattern is as follows: [KRHQSA]-[DENQ]-E-L>. Thioredoxins are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved. Bacteriophage T4 also encodes for a thioredoxin but its primary structure is not homologous to bacterial, plant and vertebrate thioredoxins. A number of eukaryotic

proteins contain domains evolutionary related to thioredoxin, all of them seem to be protein disulphide isomerases (PDI). PDI (EC 5.3.4.1) is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: - PDI major isozyme: a

5 multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deiodinase (EC 3.8. 1.4), as glutathione-insulin transhydrogenase (EC 1.8.4.2) and as a thyroid hormone-binding protein - ERp60 (ER-60; 58 Kd microsomal protein). ERp60 was originally thought to be a phosphoinositide-specific
10 phospholipase C isozyme and later to be a protease. - ERp72. - P5. All PDI contains two or three (ERp72) copies of the thioredoxin domain. The concensus pattern is as follows: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTV]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT]. The two C's form the redox-active bond.

15 Preferred polypeptides of the invention comprise the following amino acid sequence: SLHRFVLSQAKDEL (SEQ ID NO: 348), FIKFFAPWCGHCKALAPTW (SEQ ID NO: 349), and/or FIKFYAPWCGHCKTLAPTW (SEQ ID NO: 350). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the endoplasmic reticulum
20 targeting sequence domain and thioredoxin family active site domain of the sequence referenced in Table for this gene, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of this referenced sequence. The additional contiguous amino acid residues is N-terminal or C- terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain.

25 Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with
30 thioredoxin proteins. Such activities are known in the art, some of which are described elsewhere herein.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: RRG RGVPGRRRLWSAACGHCQRLQPTWN
5 DLGDKYNSMEXAKVYVAKVDCTAHS DVC SAQGV RGYPTLKLFPKPGQEAV
KYQGPRDFQTLNWMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASNFEH
VAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQHY
ELCSGNQVRGYPTLLWFRDGGKVDQYKGKRDLESLREYVESQLQRTETGA
TETVTPSEAPVLAAEPEADKGTVLALTENNFD D TIAEGITFIKFYAPWCGHC
10 KTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICKYSVRGYPTLLLFRGGK
KVSEHSGGRDLDS LHRFVLSQAKDEL (SEQ ID NO: 351). Polynucleotides
encoding these polypeptides are also provided.

This gene is expressed primarily in human chondrosarcoma and endothelial cells and to a lesser extent in a wide range of normal and diseased adult and fetal
15 tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, chondrosarcoma and other cancers and proliferative disorders.
20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, skeletal,
25 developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution in chondrosarcoma, combined with the homology to protein disulfide isomerase and phospholipase C indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

chondrosarcoma and other cancers and proliferative disorders, and possibly as a reagent for in vitro production of proteins. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the expression in endothelial cells indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:87 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2766 of SEQ ID NO:87, b is an integer of 15 to 2780, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:87, and where b is greater than or equal to a + 14.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed primarily in thyroid and thymus

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
15 not limited to, thyroid diseases including thyroid cancer and diseases of function including Grave's Disease, hyper- and hypo- thyroidism as well as Diseases of the thymus. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
20 particularly of the endocrine and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., endocrine, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene
25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in thyroid cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thyroid and thymus. Representative uses are
30 described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of

Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:88 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1047 of SEQ ID NO:88, b is an integer of 15 to 1061, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:88, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

The translation product of this gene shares sequence homology with collagen which is thought to be important as a structural material in a variety of human tissues and products including hair, nails, muscle and bone.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRPQGPAASPQRLRGLLLLLLLQLPAPSSASEIPKGKQK AHSGRGRWWTCIMECA YKGQQECLVETGALGPMAFRVHLGSQVGMDSKEK RGNV (SEQ ID NO: 352). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle and to a lesser extent in 12 week old early stage human, epididymus, healing groin wound, synovial hypoxia,

stromal cells, ulcerative colitis, breast and 8 week old embryo, as well as a variety of other normal and diseased cell types from adult and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders as well as Diseases of smooth muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 202 as residues: Glu-32 to Glu-46, Pro-63 to Ala-71, Pro-81 to Lys-90, Ser-97 to Trp-111, Lys-130 to Ser-135, Leu-147 to Cys-154, Asp-179 to Asn-186, Ser-219 to Gly-229. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in smooth muscle and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of vascular diseases and/or disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis).

atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athletes foot, and ringworm).

Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:89 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1328 of SEQ ID NO:89, b is an

integer of 15 to 1342, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:89, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

5 This gene is expressed primarily in immune cells and to a lesser extent in a wide variety of human tissues.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T cell or B cell leukemia and various immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 203 as residues: Gly-3 to Gln-9. Polynucleotides encoding said polypeptides are also provided.

 The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system diseases such as immunodeficiencies and T cell and/or B cell leukemia. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:90 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:90, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:90, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

The translation product of this gene shares sequence homology with IgE receptor. See for example, Isolation and Characterization of cDNAs coding for the Beta Subunit of the High-affinity Receptor for Immunoglobulin E, Proc. Natl. Acad. Sci. U S A. (1988 Sep.) 85(17): 6483-6487. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with IgE receptor proteins. Such activities are known in the art, some of which are described elsewhere herein. IgE and its receptors are believed to have evolved as a mechanism to protect mammals against parasites. But other and intrinsically innocuous antigens can subvert this system to provoke an allergic response. For human populations in industrialized countries, allergy and asthma now represent a far greater threat than parasitic infection, and the main impetus for current studies of the IgE system is the hope of understanding and intervening in the aetiology of allergic diseases. The high-affinity receptor for immunoglobulin (Ig) E (Fc epsilon RI) on mast cells and basophils plays a key role in IgE-mediated allergies. Fc epsilon RI is composed of one alpha, one beta, and two gamma chains, which are all required for cell surface expression of Fc epsilon RI, but only the alpha chain is involved in the binding to IgE. Fc epsilon RI-IgE interaction is highly species specific, and rodent Fc epsilon RI does not bind human IgE. New homolog can be used to develop anti-allergic agents. FcR deliver signals when they are aggregated at the cell surface. The aggregation of FcR having immunoreceptor tyrosine-based activation motifs (ITAMs) activates sequentially src family tyrosine kinases and syk family tyrosine kinases that connect transduced signals to common activation pathways shared with other receptors. FcR with ITAMs elicit cell activation, endocytosis, and phagocytosis. The nature of responses depends primarily on the cell type. The aggregation of FcR without ITAM does not trigger cell activation. Most of these FcR internalize their ligands, which can be endocytosed, phagocytosed, or transcytosed. The fate of internalized receptor-ligand complexes depends on defined sequences in the intracytoplasmic domain of the receptors. The coaggregation of different FcR results in positive or negative cooperation. Some FcR without ITAM use FcR with ITAM as signal transduction subunits. The coaggregation of antigen receptors or of FcR having ITAMs with FcR having immunoreceptor tyrosine-based inhibition motifs (ITIMs)

negatively regulates cell activation. FcR therefore appear as the subunits of multichain receptors whose constitution is not predetermined and which deliver adaptative messages as a function of the environment.

The polypeptide of this gene has been determined to have four transmembrane domains at about amino acid position 51 - 67, 89 - 105, 119 - 135, and 190 - 206 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ETRVKTSLELLRTQLEPTGTVGNTIMTSQPVPN ETIIVLPSNVINFSQAEKPEPTNQQQDSLKKHLHAEIKVIGTIQILCGMMVLSL GILASASFSPNFTQVTSTLLNSAYPFIGPFFFIISGSLSIATEKRLTKLLVHSSLV GSILSALSALVGFIILSVKQATLNPASLQCELDKNNIPTRSYVSYFYHDSLYTT DCYTAKASLAGXLSLMLICTLLEFCLAVLTAVLRWKQAYSDFPGSVLFLPH SYIGNSGMSSKMTDCGYEELLTS (SEQ ID NO: 353). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MMVLSLGILASASFSPNFTQVTSTLLNSAYPFIGPFFFI ISGSLSIATEKRLTKLLVHSSLVGSILSALSALVGFIILSVKQATLNPASLQCELDKNNIPTRSYVSYFYHDSLYTTDCYTAKASLAGXLSLMLICTLLEFCL AVLTAVLRWKQAYSDFPGSVLFLPHSYIGNSGMSSKMTDCGYEELLTS (SEQ ID NO: 354). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune system tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system diseases and/or disorders such as cancer. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in
5 certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 204 as residues: Gln-23 to Lys-39, Glu-150 to Thr-158. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells and tissues combined with the homology to IgE receptor indicates that polynucleotides and polypeptides
15 corresponding to this gene are useful for diagnosis and treatment of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell
20 lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for
25 immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity
30 disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that

influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:91 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1556 of SEQ ID NO:91, b is an integer of 15 to 1570, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:91, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 82

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GASCEGGGAAARAALGVHRSQKALLVFRRTL
30 SNLLYMPLLRGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLLSRLLHPEAH
VPPAADRGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHGN
GFPGPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSL

AFLSSLMFTSRNGLVGC GASCEGGGAAARAALGVHRSQKALLVFRRTLNL
LYMPLLRGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLRSRLLHPEAHVPP
AAD RGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHNGNFP
GPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSLA

- 5 FLSSLMFTSRNGLVGC (SEQ ID NO: 355). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

- 10 This gene is expressed primarily in activated T cells, and to a lesser extent in a wide variety of human tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 15
20
25

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 205 as residues: Pro-67 to Ser-73. Polynucleotides encoding said polypeptides are also provided.

- The tissue distribution in activated T cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immunodeficiencies. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and
- 30

elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:92 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 2936 of SEQ ID NO:92, b is an integer of 15 to 2950, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:92, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 83

The translation product of this gene was shown to have homology to the human transmembrane protein (See Genbank Accession No. gblAAC51364.11 (AF000959); all references and information available through this accession are hereby incorporated by reference herein; for example, Genomics 42 (2), 245-251
10 (1997)) which is thought to be implicated in velo-cardio-facial syndrome.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MGSAALEILGLVLCLVGWGGILACGLPMWQVTAFLD
HNIVTAQTTWKGLWMSCVVQSTGTCSAKCTTRCWL (SEQ ID NO: 356).

Polynucleotides encoding these polypeptides are also provided.

15 The gene encoding the disclosed cDNA is believed to reside on chromosome 22. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in dementia brain tissue, and to a lesser extent in a wide variety of human tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
25 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 206 as residues: Ser-201 to Tyr-217. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dementia brain tissue, combined with the homology
5 to the transmembrane protein indicates that polynucleotides and polypeptides
corresponding to this gene are useful for diagnosis and treatment of dementia, and
potentially for velo-cardio-facial syndrome. Representative uses are described in the
"Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11,
15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the
10 detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,
Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating
diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal
cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia,
mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder,
15 learning disabilities, ALS, psychoses, autism, and altered behaviors, including
disorders in feeding, sleep patterns, balance, and perception. In addition, elevated
expression of this gene product in regions of the brain indicates it plays a role in
normal neural function.

Potentially, this gene product is involved in synapse formation,
20 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or
survival. Furthermore, the protein may also be used to determine biological activity,
to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to
identify agents that modulate their interactions, in addition to its use as a nutritional
supplement. Protein, as well as, antibodies directed against the protein may show
25 utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly
available and accessible through sequence databases. Some of these sequences are
related to SEQ ID NO:93 and may have been publicly available prior to conception of
the present invention. Preferably, such related polynucleotides are specifically
30 excluded from the scope of the present invention. To list every related sequence is
cumbersome. Accordingly, preferably excluded from the present invention are one or
more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1708 of SEQ ID NO:93, b is an integer of 15 to 1722, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:93, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LKRAPPGPALAKGLLQPSSTFQALETNIGDQVR
10 RHSTAVVIREMTSYILISFVLLIGVGCIEKDQSCPVFGGRKRLHLLFVGGQLRQ
VRMLRGELSCACYRPHVQALQLGGCTCF (SEQ ID NO: 357). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the adult pulmonary system.

Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cystic fibrosis, bronchitis and any pulmonary disorders in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
20 type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, cardiovascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary surfactant, pulmonary lavage/sputum, synovial fluid and spinal fluid) or
25 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in the pulmonary system indicates that it plays a key role in the functioning of the pulmonary system. This would suggest
30 that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung and the protein product could be used either in the treatment and/or detection of these disease states. The gene

or gene product may also be useful in the treatment and/or detection of pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:94 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 621 of SEQ ID NO:94, b is an integer of 15 to 635, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:94, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene was found to be homologous to CAM proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with CAM proteins. Such activities are known in the art, some of which are described elsewhere herein.

A preferred polypeptide variant of the invention comprises the following amino acid sequence: MLCPWRTANLGLLLILTIFLVAEAGAAQPNNSLM LQTSKENHALASSSLCMDEKQITQNYSKVLAEVNTSWPVKMATNAVLC CPPIALRNLIITWEILRGQPSTCKAYKKETNETKETNCTDERITWVSRPDQ NSDLQIRTVAITHDGYRCIMVTPDGNFHRGYHLQVLVTPEVTLFQNRNRTA VCKAVAGKPAAHISWIPEGDCATKQEYWSNGTVTVKSTCHWEVHNVSTV NCHVSHLTGNKSLYIELLPVPGAKKSSKLYIPYIILTIILTIVGXIWLLKVNG CXKYKLNKPESTPVVEEDEMOPYAFYTEKNNPLXXTTNKVKASEALQSEV

DTDLHTL (SEQ ID NO:208). Polynucleotides encoding these polypeptides are also provided.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 271 - 287 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 288 to 348 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

This gene is expressed primarily in dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, auto-immunities, cancer, metastasis, wound healing, inflammation, anemias (leukemia) and other hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 208 as residues: Asp-53 to Tyr-61, Pro-105 to Ile-128, Arg-133 to Leu-140, Gln-182 to Ala-188, Pro-205 to Asn-218, Gly-259 to Ala-264, Asn-290 to Ser-302, Glu-307 to Tyr-314, Tyr-317 to Lys-332. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities.

immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:95 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general
5 formula of a-b, where a is any integer between 1 to 3784 of SEQ ID NO:95. b is an integer of 15 to 3798, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:95, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

10 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VIKLICPAAFPVYFQDMARGCVCSLCASVCIFLS
SLFPLLPSVHSVNIISCLLLSKCFEGLELMCEHL YQLSQLHVLHHIFS YLLCTP
15 (SEQ ID NO: 358). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in embryonic tissue and to a lesser extent in
in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, cancer and neoplastic states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the
25 developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, differentiating, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression
30 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in embryonic tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental anomalies, fetal deficiencies and pre-natal disorders, as well as abnormal cell proliferation and/or differentiation, neoplastic states and cancer.

5 Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental
10 tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent
15 of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating,
20 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in
25 proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,
30 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:96 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically
5 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2669 of SEQ ID NO:96, b is an integer of 15 to 2683, where both a and b correspond to the positions of nucleotide
10 residues shown in SEQ ID NO:96, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

The translation product of this gene shares sequence homology with inter-alpha-trypsin inhibitor which is thought to be important in inhibition of trypsin and
15 other serine proteases (See Genbank Accession No. pirlS30350IS30350; all references and information available through this accession are hereby incorporated herein by reference; for example, Eur. J. Biochem. 179 (1), 147-154 (1989), J. Biol. Chem. 264 (27), 15975-15981 (1989), and J. Biol. Chem. 266 (2), 747-751 (1991)).

Contact of cells with supernatant expressing the product of this gene has been
20 shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are
25 not limited to, activating monocytes, and to a lesser extent, other immune and/or hematopoietic cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

30 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the

following amino acid sequence: YXIPGSTHASGRQGRGSGREDDSGPPPSTVINQ
NETFANIIFKPTVVQQARIAQNGILGDFIIRYDVNREQSIGDIQVLNGYFVHYF
APKDLPLPKNVVFLDSSASMVGTKLRQTKDALFTILHDLRPQDRFSIIGFS
NRIKVWKDHLISVTPDSIRDGKVYIHHMSPTGGTDINGVLQRAIRLLNKYVAH
5 SGIGDRSVSLIVFLTDG KPTVGETHTLKILNNTREAARGQVCIFTIGIGNDVD
FRLLEKLSLENCGLTRRVHEEEDAGSQLIGFYDEIRTPLLSDIRIDYPPSSVVQ
ATKTLFPNYFNGSEIIIAGKLVDRLDHLHVEVTASNSKKFIILKTDVPRPQK
AGKDVGTGSPRPGDGEGDXNHIERLWSYLTTKELLSSWLQSDDEPEKERLRQ
RAQALAVSYRFLTPFTSMKLRGPVPRMDGLEEAHGMSAAMGPEPVVQSVR
10 GAGTQPGPLLKKPYQPRIKISKTSVDGDPHFVVDPLSRLTVCFNIDGQPGDIL
RLVSDHRDSGVTVNGELIGAPAPPNGHKKQRTYLRITILINKPERSYLEITPS
RVILDGGDRLVLPCNQSVVVGSGWGLEVSVSANANVTVTIQGSIAFVILHLYK
KPAPFQRHHLGFYIANSEGLSSNCHGLLGQFLNQDARLTEDPAGPSQNLTHP
LLLQVGEGPEAVLTVKGHQVPVWVKQRKIYN GEEQXDCWFARNMPPN
15 (SEQ ID NO: 359). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta and adipose tissue and to a lesser extent in several other organs and tissues including cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of developing organs and metabolic diseases, in addition to vascular diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
25 tissues or cells, particularly of the developing systems and metabolic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative
30 to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 210 as residues: Lys-5 to Lys-10, Asn-33 to Lys-39, Asp-48 to Lys-54, Pro-62 to Asp-67, Asn-116 to Arg-123, His-157 to Ala-162, Val-242 to Lys-249, Val-251 to Asp-264. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in placenta, combined with the homology to inter-alpha-trypsin inhibitor and the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of developing and metabolic systems. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease,

vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Polynucleotides and polypeptides of the invention are also useful for the treatment, detection, and/or prevention of inflammation, tumor invasion and metastasis, wound healing, liver disease, 5 disseminated intravascular coagulation, alzheimer's Disease, ophthalmic disease, apoptosis, tissue remodeling, intrauterine growth retardation, preeclampsia, angiogenesis, cell migration, fetal development, trophoblast implantation, ovulation, pemphigus and psoriasis, and antiviral therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate 10 cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly 15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:97 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is 20 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2167 of SEQ ID NO:97, b is an integer of 15 to 2181, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:97, and where b is greater than or equal to a + 14.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 88

The translation product of this gene was shown to have homology to the human colon carcinoma antigen NY-CO-7 (See Genbank and Geneseq Accession Nos. gb|AAC18038.11 (AF039689) and WO9904265; all references available through this accession are hereby incorporated herein by reference; for example, Int. J. Cancer 30 76 (5), 652-658 (1998)).

This gene is expressed primarily in breast and breast cancer and to a lesser extent in several other organs and tissues including cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of reproductive organs and the gastrointestinal system, including cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, breast milk, chyme, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 211 as residues: Gly-22 to Gly-28, Leu-71 to Phe-77, Asn-101 to Val-108, Pro-122 to Ser-127, Arg-149 to Pro-154, Gly-191 to Phe-196, Pro-199 to Thr-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in breast and breast cancer tissue, combined with the homology to a colon cancer antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the reproductive systems and cancers. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:98 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1943 of SEQ ID NO:98, b is an integer of 15 to 1957, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:98, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

The translation product of this gene shares sequence homology with the amino acid and protein sequence of a *Xenopus* transmembrane protein of unknown function. The very 5'-end of the contig is identical to the mRNA for the human LGN mosaic protein. Based on the sequence similarity, the translation product of this gene is

expected to share at least some biological activities with LGN mosaic proteins. Such activities are known in the art, some of which are described elsewhere herein.

Preferred polypeptides of the invention comprise the following amino acid sequence:

PRVRPPTKALAVTFTTFVTEPLKHIGKGTGEFIKALMKEIPALLHLPVLIIMAL

5 AILSFCYGAGKSVHVLRHIGGPEREPPQALRPDRRRQEEIDYRPDGGAGDAD
FHYRGQMGPTTEQGPYAKTYEGRREILRERDVDLRFQTGNKSPEVLRAFDVPD
AEAREHPTVVPSHKSPVLDTKPKETGGILGEGTPKESSTESSQSAKPVSGQDTS
GNTEGSPA AEKAQLKSEAAGSPDQGSTYSPARGVAGPRGQDPVSSPCG (SEQ
ID NO:339). Polynucleotides encoding such polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in small intestine and adipocytes and to a lesser extent in various other normal and transformed cell types, mostly of endocrine
15 origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, conditions of growth and metabolism. Similarly, polypeptides and
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, gastrointestinal, and
25 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, bile, chyme, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 212 as residues: Pro-40 to Gly-68, Gly-79 to Arg-93, Phe-106 to Glu-114, Pro-122 to His-129, Thr-143 to Gly-149, Gly-155 to Ala-168,

Val-171 to Gly-182, Ala-195 to Pro-207, Pro-214 to Val-220. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in small intestine indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of disorders of growth and metabolism as well as endocrine abnormalities. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:99 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1098 of SEQ ID NO:99, b is an integer of 15 to 1112, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:99, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

The translation product of this gene shares sequence homology with IgE receptor beta chain which is thought to be important in immune function.

This gene is expressed primarily in kidney medulla tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and renal diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and renal systems,

expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, renal, urogenital, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney renal medulla tissue, combined with the homology to the IgE receptor beta chain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of immune and renal disorders. The protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Alternatively, this gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the

protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:100 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 873 of SEQ ID NO:100, b is an integer of 15 to 887, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:100, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 91

The translation product of this gene shares sequence homology with Diff 40 gene product (See Genbank Accession No. gblAAC51134.1; all references and information available through this reference are hereby incorporated herein).

Preferred polypeptides of the invention comprise the following amino acid sequence: PRVRSIKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDIN
DLGTIKLSLEVTWSPFDKDDQPSAASSVNKASTVTKRFSTYSQSPDTPS
LREQAFYNMLRRQEELENGTAWSLSSSESSDDSSSPQLSGTARHSPAPRPLV
QQPEPLPIQVAFRRPETPSSGPLDEEGAVAPVLANGHAPYSRTLHISEASVNA
ALAEASVEAVGPKSLSWGSPPTHPAPTHGKHPSVPVPPALDPGHSATST
LGTGTSVPTSTDPAPSAHLDSVHKSTDGSELPGPHTTTTGSTYSAITTTHS
APSPLTHTTTGSTHKPIISTLTTTGPTLNIIGPVQTTTSPHTTMPSPSSHNSPQ
YVDFCSSVCDNIFVHYVIGIFFHTLYSSKTL (SEQ ID NO:360), and/or PRVRS
IKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDINDLGTIKLSLEVTWSP
FDKDDQPSAASSVNKASTVTKRFSTYSQSPDTPSLREQAFYNMLRRQEELE
NGTAWSLSSSESSDDSSSPQLSGTARHSPAP RPLVQQPEPLPIQVAFRRPET

PSSGPLDEEGAVAPVLANGHAPYSRTLSEASVNAALAEASVEAVGPKSL
SWGSPPTHAPATHGKHPSPVPPALDPGHSATSSTLGTTGSVPTSTD (SEQ ID
NO: 361). Polynucleotides encoding these polypeptides are also provided.

Polypeptides of the invention do not consist of the primary amino acid sequence
5 shown as Geneseq Accession No.W69430, which is hereby incorporated herein by
reference.

This gene is expressed primarily in liver and to a lesser extent in gall bladder
tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, metabolic and endocrine diseases and/or disorders, particularly hepatic
and gall bladder disorders. Similarly, polypeptides and antibodies directed to these
polypeptides are useful in providing immunological probes for differential
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the metabolic and endocrine systems, expression of this
gene at significantly higher or lower levels is routinely detected in certain tissues or
cell types (e.g., hepatic, metabolic, gall bladder, gastrointestinal, and cancerous and
wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, bile, synovial
20 fluid and spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression
level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic
epitopes shown in SEQ ID NO: 214 as residues: Val-9 to Cys-14, Pro-42 to Thr-47,
25 Thr-56 to Ala-64, Asp-88 to His-98, Cys-128 to Ser-136, Arg-153 to Trp-161.
Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in liver and gall bladder, combined with the homology
to the diff 40 gene product indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the study and treatment of endocrine and
30 metabolic disorders. polynucleotides and polypeptides corresponding to this gene are
useful for the detection and treatment of liver disorders and cancers. Representative
uses are described in the "Hyperproliferative Disorders", "infectious disease", and

"Binding Activity" sections below, in Example 11, and 27, and elsewhere herein.

Briefly, the protein can be used for the detection, treatment, and/or prevention of hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:101 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1234 of SEQ ID NO:101, b is an integer of 15 to 1248, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:101, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 3 - 19 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

This gene is expressed primarily in fetal brain and to a lesser extent in pancreas tumor, melanocyte and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly neurodevelopmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution in fetal brain tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders of the central nervous system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not
15 limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive
20 disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,
25 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show
30 utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:102 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or
 5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1827 of SEQ ID NO:102, b is an integer of 15 to 1841, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:102, and where b is greater than or equal to a + 14.

10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 93**

The translation product of this gene shares sequence homology with a probable membrane protein YGL054c -yeast (*Saccharomyces cerevisiae*). Moreover,

The translation product of this gene also have homology to the human and mouse cornichon protein which is known to be necessary for both anterior-posterior
 15 and dorsal-ventral pattern formation in conjunction with the EGF receptor signaling process (See Genbank Accession Nos. gblAAC98388.11 (AF104398), and splP52159; all references and information available through these accessions are hereby incorporated herein by reference; for example, Cell 81 (6), 967-978 (1995)).

The polypeptide of this gene has been determined to have two transmembrane
 20 domains at about amino acid position 57 - 73, and 121 - 137 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 - 14 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

25 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: YGCEKTTEGRRRRRRMEAVVFVFSLLDCCAL
 IFLSVYFIITLSDLECDYINARSCSKLNKWWIPELIGHTIVTVLLLSLHWF
 30 IFLNLPLVATWNIYRYIMVPSGNMGVFDPTIEHNRGQLKSHMKEAMIKLGFH
 LLCFFMYLYSMILALIND (SEQ ID NO:362). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endometrial tumor, T cell helper II cells, microvascular endothelial cells, Raji cells treated with cyclohexamide and umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
10 not limited to, immune, hematopoietic, and vascular diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is
15 routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
20 individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 216 as residues: Ser-39 to Asn-45, Asn-103 to Ser-109. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in activated T-cells indicates that polynucleotides and
25 polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders involving activated T-cells. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or
30 activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:103 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 671 of SEQ ID NO:103, b is an integer of 15 to 685, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:103, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARAPAPSLPPLPSPAPALAPAHSLGLLLGRMS
10 GSSLPSALALSLLLVSGLLPGPAAQNVVRVQSGQDQ (SEQ ID NO: 363).
Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in dendritic cells and to a lesser extent in healing abdomen wound, and pancreas islet cell tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly wound healing disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
20 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or
25 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 217 as residues: Gln-34 to Lys-40. Polynucleotides
30 encoding said polypeptides are also provided.

The tissue distribution in dendritic cells and early healing wound indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating

wounds to enhance the healing process. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or
5 activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene
10 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and
15 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits
20 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their
25 interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are
30 related to SEQ ID NO:104 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is

cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1154 of SEQ ID NO:104, b is an integer of 15 to 1168, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:104, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of aortic smooth muscle cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both smooth muscle cells, and in other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating smooth muscle cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

This gene is expressed primarily in pancreatic carcinoma, gall bladder and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic and immune diseases and/or disorders, particularly cancers, such as pancreatic carcinoma and gall bladder tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such

a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 218 as residues: Lys-34 to Ile-41. Polynucleotides
5 encoding said polypeptides are also provided.

The tissue distribution in pancreatic carcinoma and gall bladder indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancer, such as pancreatic carcinoma and gall bladder tumors.

Representative uses are described here and elsewhere herein. Alternatively, the
10 detected calcium flux biological activity indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to
15 determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

20 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:105 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
25 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1161 of SEQ ID NO:105, b is an integer of 15 to 1175, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:105, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 10 - 26 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 27 to 48 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ib membrane proteins.

This gene is expressed primarily in osteosarcoma, wilm's tumor, ovarian cancer and in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammatory diseases and cancers, such as osteosarcoma, wilm's tumor and ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, renal, reproductive, immune, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 219 as residues: Ser-30 to Pro-35. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory conditions and cancer, such as osteosarcoma, wilm's tumor and ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases

and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

5 Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:106 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1007 of SEQ ID NO:106, b is an

integer of 15 to 1021, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:106, and where b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

5 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSKDCVLYAFLDPGMAVPLFLYIFTLLPLLPLFLLSLCFSP LTVKRSSSSSESKSSL (SEQ ID NO: 364). Polynucleotides encoding
10 these polypeptides are also provided.

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
15 not limited to, ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types
20 (e.g., reproductive, ovarian, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial amniotic fluid, fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 220 as residues: Thr-28 to Ser-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in ovarian tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing cancer,
30 e.g., ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of

developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:107 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 816 of SEQ ID NO:107, b is an integer of 15 to 830, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:107, and where b is greater than or equal to a + 14.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

This gene is expressed primarily in macrophages and breast cancer tissue and to a lesser extent in osteoblasts and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system dysfunction; inflammation; breast cancer; cancer; osteoporosis; osteopetrosis; peristaltic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 221 as residues: Glu-16 to Ala-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. Expression in macrophages and other hematopoietic cell types indicates that this gene product is involved in the regulation of hematopoietic cell survival, proliferation, differentiation, or activation. It is involved in the control of such processes as immune surveillance, antigen presentation, T cell activation, cytokine release, and inflammation. Expression in breast cancer tissue may possibly correlate with the diagnosis and differentiation of cancerous tissue from normal breast tissue.

Expression in osteoblasts and osteoclasts may implicate this gene product in the process of bone turnover, and target it as a likely candidate for the treatment of osteoporosis and/or osteopetrosis. Finally, expression in smooth muscle may indicate an involvement in the normal function of numerous internal organs and in the function of the digestive system.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:108 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1287 of SEQ ID NO:108, b is an integer of 15 to 1301, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:108, and where b is greater than or equal to a + 14.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	IIDPTK41	209965 06/11/98	pCMV'Sport 3.0	11	1564	1	1564	39	39	124	1	26	27	369
2	HFXGT26	209965 06/11/98	Lambda ZAP II	12	1757	1	1757	13	13	125	1	22	23	85
3	HLTHGX30	209965 06/11/98	Uni-ZAP XR	13	1373	1	1373	13	13	126	1	41	42	43
4	HLTHIG37	209965 06/11/98	Uni-ZAP XR	14	3740	1908	3740	50	50	127	1	1	2	319
4	HLTHIG37	209965 06/11/98	Uni-ZAP XR	109	1932	98	1932	313	313	222	1	35	36	42
5	HINTNIZ90	209965 06/11/98	pSport1	15	1196	1	1196	282	282	128	1	21	22	45

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
6	HP1BX03	209965 06/11/98	Uni-ZAP XR	16	2209	1	2178	81	81	129	1	29	30	709
7	H6EDY30	209965 06/11/98	Uni-ZAP XR	17	1774	1	1774	321	321	130	1	29	30	414
8	HAMGR28	209965 06/11/98	pCMV Sport 3.0	18	1674	47	1674	98	98	131	1	18	19	242
8	HAMGR28	209965 06/11/98	pCMV Sport 3.0	110	1534	1	1534	40	40	223	1	18	19	203
9	HAPNZ94	209965 06/11/98	Uni-ZAP XR	19	2018	255	2018	287	287	132	1	36	37	312
10	HATCP77	209965 06/11/98	Uni-ZAP XR	20	2098	1	2098	37	37	133	1	21	22	182

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
11	HDABR72	209965 06/11/98	pSportI	21	1746	1	1746	28	28	134	1	29	30	146
12	HDPKB18	209965 06/11/98	pCMVSPORT 3.0	22	2876	1	2876	98	98	135	1	21	22	122
12	HDPKB18	209965 06/11/98	pCMVSPORT 3.0	111	2871	1	2871	87	87	224	1	21	22	42
13	HEQCC55	209965 06/11/98	pCMVSPORT 3.0	23	1052	30	1052	62	62	136	1	27	28	112
13	HEQCC55	209965 06/11/98	pCMVSPORT 3.0	112	1037	1	1037	57	57	225	1	27	28	155
14	HETDE26	209965 06/11/98	Uni-ZAP XR	24	1541	1	1541	205	205	137	1	29	30	139

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
15	HOEDI184	209965 06/11/98	Uni-ZAP XR	25	2079	1	2079	256	256	138	1	20	21	404
16	HPIBT55	209965 06/11/98	Uni-ZAP XR	26	1947	129	1947	253	253	139	1	30	31	95
17	HSLCS05	209965 06/11/98	Uni-ZAP XR	27	3379	1	3354	168	168	140	1	23	24	239
18	HDPDD03	209965 06/11/98	pCMVSPORT 3.0	28	2006	1	2006	233	233	141	1	21	22	53
19	HDPDI66	209965 06/11/98	pCMVSPORT 3.0	29	3070	1	3070	93	93	142	1	45	46	66
20	HDTDQ23	209965 06/11/98	pCMVSPORT 2.0	30	2227	1	2206	148	148	143	1	20	21	108

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
20	IIDTDQ23	209965 06/11/98	pCMVSPORT 2.0	113	2214	1	2206	148	226	1	20	21	73
21	HE2PY40	209965 06/11/98	Uni-ZAP XR	31	1288	1	1288	147	144	1	22	23	83
22	HEONM66	209965 06/11/98	pSport1	32	3280	1	3280	89	145	1	24	25	166
22	HEONM66	209965 06/11/98	pSport1	114	3300	1	3300	98	227	1	20	21	166
23	HKAEG43	209965 06/11/98	pCMVSPORT 2.0	33	1297	1	1297	32	146	1	29	30	70
23	HKAEG43	209965 06/11/98	pCMVSPORT 2.0	115	1286	1	1286	21	228	1	29	30	70

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
24	HLHDP65	209965 06/11/98	Uni-ZAP XR	34	2184	1	2184	19	19	147	1	19	20	412
24	HLHDP65	209965 06/11/98	Uni-ZAP XR	116	2189	1	2189	26	26	229	1	21	22	272
25	HLMD003	209965 06/11/98	Uni-ZAP XR	35	949	1	949	72	72	148	1	45	46	84
26	HIMAGK93	209965 06/11/98	Uni-ZAP XR	36	3338	162	1884	164	164	149	1	30	31	153
27	HIMEAL02	209965 06/11/98	Lambda ZAP II	37	1563	1	1563	237	237	150	1	33	34	129
28	HMKCH52	209965 06/11/98	pSport1	38	1048	1	1048	53	53	151	1	17	18	61

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
29	HCEFB69	209965 06/11/98	Uni-ZAP XR	39	1430	1	1430	188	188	152	1	24	25	224
30	HINFEC43	203027 06/26/98	Uni-ZAP XR	40	2103	209	2058	488	488	153	1	15	16	68
31	HSPMG77	203027 06/26/98	pSport1	41	2349	1	2349	130	130	154	1	46	47	83
32	HISQAC69	203027 06/26/98	Uni-ZAP XR	42	1559	1	1559	146	146	155	1	21	22	60
33	HISTB186	203027 06/26/98	Uni-ZAP XR	43	1766	1	1766	120	120	156	1	24	25	83
34	HILDQR62	203027 06/26/98	pCMVSPORT 3.0	44	2572	427	2572	520	520	157	1	18	19	161

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
35	HUVDJ43	203027 06/26/98	Uni-ZAP XR	45	526	69	526	89	89	158	1	31	32	146
36	HADCP14	203027 06/26/98	pSport1	46	1032	1	1032	35	35	159	1	20	21	142
37	HBXCF95	203027 06/26/98	ZAP Express	47	2680	1	2680	118	118	160	1	22	23	50
38	HEQBU15	203027 06/26/98	pCMVSPORT 3.0	48	1730	1	1730	56	56	161	1	26	27	64
39	HL1BD22	203027 06/26/98	Uni-ZAP XR	49	1275	1	1275	53	53	162	1	39	40	58
40	HOEEU24	203027 06/26/98	Uni-ZAP XR	50	1762	1	1762	113	113	163	1	21	22	374

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HOEEU24	203027 06/26/98	Uni-ZAP XR	117	1763	1	1763	113	113	230	1	21	22	81
41	HTTBR96	203027 06/26/98	Uni-ZAP XR	51	2059	1	2059	96	96	164	1	26	27	63
42	HWHIQS55	203027 06/26/98	pCMVSPORT 3.0	52	3282	1	3282	169	169	165	1	26	27	742
43	HCEEK50	203027 06/26/98	Uni-ZAP XR	53	1860	1	1860	233	233	166	1	17	18	213
44	HCWBU94	203027 06/26/98	ZAP Express	54	770	1	770	109	109	167	1	26	27	212
45	HE2NR62	203027 06/26/98	Uni-ZAP XR	55	1093	1	1093	145	145	168	1	38	39	74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
46	HHSGH19	203027 06/26/98	Uni-ZAP XR	56	632	1	632	291	291	169	1	15	16	47
47	HDPGT01	203027 06/26/98	pCMVSPORT 3.0	57	2687	138	2687	8	8	170	1	28	29	87
48	HOBAF11	203027 06/26/98	pBluescript	58	619	153	579	166	166	171	1	30	31	41
49	HOHCA35	203027 06/26/98	pCMVSPORT 2.0	59	1378	1	1378	153	153	172	1	15	16	47
50	HPMGP24	203027 06/26/98	Uni-ZAP XR	60	1126	1	1126	215	215	173	1	33	34	232
51	HSDIE16	203027 06/26/98	Uni-ZAP XR	61	2078	1	2078	182	182	174	1	29	30	44

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Secreted Portion	Last AA of ORF
52	HSOBK48	203027 06/26/98	Uni-ZAP XR	62	762	1	762	433	433	175	1	16	84
53	HTAD1139	203027 06/26/98	Uni-ZAP XR	63	1094	1	1094	173	173	176	1	24	65
54	HTUSGT36	203027 06/26/98	pSport1	64	1361	1	1361	112	112	177	1	16	54
55	HVA AE95	203027 06/26/98	pSport1	65	947	1	947	325	325	178	1	14	82
56	HIHEAI125	203071 07/27/98	pCMVSPORT 3.0	66	1376	1	1376	43	43	179	1	31	330
56	HIHEAI125	203071 07/27/98	pCMVSPORT 3.0	118	1375	1	1375	43	43	231	1	31	71

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
57	IIBJY92	203071 07/27/98	Uni-ZAP XR	67	2434	487	2366	548	548	180	1	29	30	40
58	HICLW50	203071 07/27/98	Lambda ZAP II	68	1086	1	1086	255	255	181	1	17	18	51
59	HDRMF68	203071 07/27/98	pSport1	69	1262	1	1262	309	309	182	1	22	23	54
60	HOUGG12	203071 07/27/98	Uni-ZAP XR	70	1642	35	1642	116	116	183	1	22	23	61
61	HEEAQ11	203071 07/27/98	Uni-ZAP XR	71	921	1	921	213	213	184	1	28	29	147
62	HEEAZ65	203071 07/27/98	Uni-ZAP XR	72	906	1	906	182	182	185	1	19	20	160

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HIEGAN94	203071 07/27/98	Uni-ZAP XR	73	680	1	680	133	186	1	23	24	121
64	HIFXBL33	203071 07/27/98	Lambda ZAP II	74	1633	1	1633	152	187	1	24	25	162
65	HILIBD68	203071 07/27/98	pCMVSPORT 1	75	1022	1	1022	186	188	1	35	36	50
66	HILTCO33	203071 07/27/98	Uni-ZAP XR	76	1184	1	1184	80	189	1	18	19	64
67	HILYAC95	203071 07/27/98	pSport1	77	312	1	312	92	190	1	16	17	46
68	HNFGEF20	203071 07/27/98	Uni-ZAP XR	78	1370	38	1370	206	191	1	45	46	143

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Secreted Portion	Last AA of ORF
69	HNHKS18	203071 07/27/98	Uni-ZAP XR	79	368	1	368	125	125	192	1	36	81
70	HSLJW78	203071 07/27/98	Uni-ZAP XR	80	1088	1	1088	159	159	193	1	20	44
71	HIIFHD01	203071 07/27/98	Uni-ZAP XR	81	1862	1	1862	177	177	194	1	16	41
72	HILWAE11	203071 07/27/98	pCMVSPORT 3.0	82	1618	1	1618	85	85	195	1	27	259
73	HICYBN55	203071 07/27/98	pBluescript SK-	83	2034	1	1934	341	341	196	1	19	117
73	HICYBN55	203071 07/27/98	pBluescript SK-	119	1022	78	1022		3	232	1	1	225

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	S' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
74	HEONX38	203071 07/27/98	pSport1	84	2240	5	2240	23	23	197	1	23	24	698
74	HEONX38	203071 07/27/98	pSport1	120	2311	1	2311	24	24	233	1	23	24	314
75	HLDQU79	203071 07/27/98	pCMVSPORT 3.0	85	1488	1	1488	99	99	198	1	23	24	348
76	HSYBK21	203071 07/27/98	pCMVSPORT 3.0	86	3174	1	1466	119	119	199	1	29	30	401
77	HELCB12	203071 07/27/98	Uni-ZAP XR	87	2780	2110	2738	120	120	200	1	30	31	324
78	HTHDS25	203071 07/27/98	Uni-ZAP XR	88	1061	1	1061	70	70	201	1	15	16	90

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
79	HFHHO70	203071 07/27/98	pSport1	89	1342	1	1271	141	141	202	1	30	31	243
79	HFHHO70	203071 07/27/98	pSport1	121	1286	1	1279	131	131	234	1	30	31	93
80	HPMEI86	203071 07/27/98	Uni-ZAP XR	90	770	40	770	50	50	203	1	30	31	75
81	HISOBV29	203071 07/27/98	Uni-ZAP XR	91	1570	207	1570	244	244	204	1	24	25	248
82	HWABY10	203071 07/27/98	pCMVSPORT 3.0	92	2950	78	2914	263	263	205	1	22	23	168
83	HACCI17	203071 07/27/98	Uni-ZAP XR	93	1722	336	1714	461	461	206	1	24	25	218

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HACCI17	203071 07/27/98	Uni-ZAP XR	122	1380	12	1380	135	135	235	1	24	25	72
84	HAPQT22	203070 07/27/98	Uni-ZAP XR	94	635	1	635	132	132	207	1	17	18	72
85	HIDPBO81	203070 07/27/98	pCMVSPORT 3.0	95	3798	1	3798	265	265	208	1	26	27	348
85	HIDPBO81	203070 07/27/98	pCMVSPORT 3.0	123	3793	1	3793	255	255	236	1	26	27	348
86	HIDPGI49	203070 07/27/98	pCMVSPORT 3.0	96	2683	1	2640	266	266	209	1	29	30	72
87	HIDTBV77	203070 07/27/98	pCMVSPORT 2.0	97	2181	1	2181	326	326	210	1	22	23	608

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
88	HFIUE82	203070 07/27/98	pSport1	98	1957	1	1957	24	24	211	1	23	24	251
89	HHEND31	203070 07/27/98	pCMVSPORT 3.0	99	1112	1	1112	109	109	212	1	25	26	225
90	HKMND01	203069 07/27/98	pBluescript	100	887	1	887	23	23	213	1	26	27	50
91	HLDBI84	203069 07/27/98	pCMVSPORT 3.0	101	1248	1	1248	50	50	214	1	35	36	171
92	HLTEK17	203069 07/27/98	Uni-ZAP XR	102	1841	1	1841	112	112	215	1	13	14	47
93	HEBEJ18	203069 07/27/98	Uni-ZAP XR	103	685	7	649	51	51	216	1	15	16	139

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
94	HMEAI48	203069 07/27/98	Lambda ZAP II	104	1168	1	1168	95	95	217	1	29	30	40
95	HNIGN91	203069 07/27/98	Uni-ZAP XR	105	1175	161	1175	184	184	218	1	24	25	51
96	HODAE92	203069 07/27/98	Uni-ZAP XR	106	1021	1	1021	123	123	219	1	29	30	48
97	HODDF13	203069 07/27/98	Uni-ZAP XR	107	830	1	830	46	46	220	1	27	28	41
98	HCDCF30	203027 06/26/98	Uni-ZAP XR	108	1301	102	1301	151	151	221	1	14	15	40

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization

probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

5 The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

10 The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during
15 recombinant production.

 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40
20 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

25 Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the
30 information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of

these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95%

"identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence

that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid.

These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

- 5 As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present
10 invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity.
- 15 Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.
- 20 If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the
25 query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from
30 the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and

C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

5 For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is
10 subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no
15 residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual
20 corrections are to be made for the purposes of the present invention.

 The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced
25 by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E.
30 coli).

 Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II. Lewin. B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

5 Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988
10 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

15 Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1 α . They used random mutagenesis to generate over 3,500 individual IL-1 α mutants that averaged 2.5 amino acid changes per variant over
20 the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in
25 activity from wild-type.

 Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted
30 form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic

activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, 5 inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

10 The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these 15 positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. 20 For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are 25 surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions 30 involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of

the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of the present invention having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polypeptide to have an amino acid sequence which comprises the amino acid sequence of the present invention, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of the present invention or fragments thereof (e.g., the mature form and/or other fragments described herein), is

1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

Polynucleotide and Polypeptide Fragments

5 In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt
10 in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

15 Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-
20 1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini.
25 Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the
30 deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the

invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these

fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

5 **Fusion Proteins**

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second
10 protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous
15 functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the
20 polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-
25 polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the
30

IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion
5 proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified,
10 would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol.
15 Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue,
20 Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767
25 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

30 The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral

vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

5 The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

10 The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding
15 portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin
20 resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as
25 CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-
9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and
30 ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1

and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods
5 are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

10 A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most
15 preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical
20 synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also
25 include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some
30 proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or replace endogenous genetic material (e.g., coding
5 sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with the polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter and/or enhancer) and endogenous polynucleotide sequences
10 via homologous recombination (see, e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989); and Zijlstra et al., Nature 342:435-438 (1989), the disclosures of each of which are incorporated by reference in their
15 entireties).

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as
20 reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat
25 polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted
30 exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids

containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or

translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more
5 restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

10 The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a
15 unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological
20 samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with
25 one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a
30 particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the

present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as
5 molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an
10 immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

15 A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene
20 expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (^{125}I , ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^3H), indium (^{112}In), and technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and
25 biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which
30 emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as

deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ^{131}I , ^{112}In , $^{99\text{m}}\text{Tc}$), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of $^{99\text{m}}\text{Tc}$. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic

anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus. Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, 5 Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide 10 or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ 15 rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of 20 T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic 25 and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or 30 IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response.

Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Bimaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Nocardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae,

Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related
5 infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin
10 diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not
15 limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g.,
20 dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present
25 invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against
30 infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vasculature (including vascular and lymphatics), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable

of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell
5 membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

10 The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations,
15 polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

20 Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers.
25 The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

30 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a

candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3'

5 Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3'

10 Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

20 Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

30 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only
5 A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said
10 cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the
15 ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide
20 sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human
25 cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological
30 sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X

wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one
5 nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid
10 molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA
15 molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence
20 selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

25 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

30 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a

biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid
5 sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

10 Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

15 Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

25 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

30 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1

and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is

defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

25	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited</u>
	<u>Plasmid</u>	
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
	Zap Express	pBK
30	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSport 2.0	pCMVSport 2.0

pCMVSPORT 3.0

pCMVSPORT 3.0

pCR[®]2.1pCR[®]2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSPORT1, pCMVSPORT 2.0 and pCMVSPORT 3.0. were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lacmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional

plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P - γ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μl of reaction mixture with 0.5 μg of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl_2 , 0.01% (w/v) gelatin, 20 μM each of dATP, dCTP, dGTP, dTTP, 25

pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of

the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

5 A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

10 Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After
15 labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

 Multiple Tissue Northern (MTN) blots containing various human tissues (H)
20 or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed
25 according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

 An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of
30 conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing

individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers
10 used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a
15 bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at
20 the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and
25 confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG
30 (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1

mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The

origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without
5 mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive
10 Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4
15 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using
20 a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the
25 above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus**Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

5 The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by
10 gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method
15 described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the
20 transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

25 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell
30 culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then

resuspended in a microcentrifuge tube containing 200 μ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

5 To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and
10 cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 μ Ci of 35 S-methionine and 5 μ Ci 35 S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

15 Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian
20 cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient
25 transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include,
30 for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109),

pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

5 Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

 The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing
10 cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker
15 is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the
20 production of proteins.

 Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the
25 CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

30 Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be
5 modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

10 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

15 Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five μ g of the expression plasmid pC6 is cotransfected with 0.5 μ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded
20 in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM,
25 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 -
30 PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

30

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGC
CCAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAAA
CCCAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGT
GGTGGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGG
5 ACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTA
CAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACT
GGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCA
ACCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAAC
CACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAG
10 GTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGT
GGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCT
CCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTG
GACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCA
TGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGG
15 GTAAATGAGTGCGACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

25 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in

any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

5 The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as
10 described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

 Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method
15 makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody
20 whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

 It will be appreciated that Fab and F(ab')₂ and other fragments of the
25 antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic
30 chemistry.

 For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced

using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening

Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45

minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl₂ (anhyd); 0.00130 mg/L CuSO₄·5H₂O; 0.050 mg/L of Fe(NO₃)₃·9H₂O; 0.417 mg/L of FeSO₄·7H₂O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl₂; 48.84 mg/L of MgSO₄; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO₃; 62.50 mg/L of NaH₂PO₄·H₂O; 71.02 mg/L of Na₂HPO₄; 4320 mg/L of ZnSO₄·7H₂O; .002 mg/L of Arachidonic Acid; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319

mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of

5 Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for

10 endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

15 On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the

20 polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

25 **Example 12: Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The

30 binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six

members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at
5 higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2,
10 Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two
15 groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

20 Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the
25 proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>		<u>JAKs</u>				<u>STATs</u>	<u>GAS(elements) or ISRE</u>
		<u>tyk2</u>	<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>			
	<u>IFN family</u>							
5	IFN-a/B	+	+	-	-	1,2,3	ISRE	
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)	
	Il-10	+	?	?	-	1,3		
	<u>gp130 family</u>							
10	IL-6 (Pleiotrophic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)	
	Il-11(Pleiotrophic)	?	+	?	?	1,3		
	OnM(Pleiotrophic)	?	+	+	?	1,3		
	LIF(Pleiotrophic)	?	+	+	?	1,3		
	CNTF(Pleiotrophic)	-/+	+	+	?	1,3		
15	G-CSF(Pleiotrophic)	?	+	?	?	1,3		
	IL-12(Pleiotrophic)	+	-	+	+	1,3		
	<u>g-C family</u>							
	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS	
20	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)	
	IL-7 (lymphocytes)	-	+	-	+	5	GAS	
	IL-9 (lymphocytes)	-	+	-	+	5	GAS	
	IL-13 (lymphocyte)	-	+	?	?	6	GAS	
	IL-15	?	+	?	+	5	GAS	
25								
	<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)	
	IL-5 (myeloid)	-	-	+	-	5	GAS	
	GM-CSF (myeloid)	-	-	+	-	5	GAS	
30								
	<u>Growth hormone family</u>							
	GH	?	-	+	-	5		

PRL	?	+/-	+	-	1,3,5
EPO	?	-	+	-	5

GAS(B-CAS>IRF1=IFP>>Ly6)

Receptor Tyrosine Kinases

5	EGF	?	+	+	-	1,3
	PDGF	?	+	+	-	1,3
	CSF-1	?	+	+	-	1,3

GAS (IRF1)

GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

10 5':GCGCCTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

15 PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

20 5':CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATGATTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCGCCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCT
25 AGGCTTTTGCAAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter
30 molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not
5 contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance
10 gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS
15 with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-
20 2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

25 The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells
30 (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately
5 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells
10 containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

15 During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

20 The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and
25 resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells
30 into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

The above protocol may be used in the generation of both transient, as well as, stable transfected cells, which would be apparent to those of skill in the art.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing

10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM
5 KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400
10 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in
15 the 96-well plate (or 1×10^5 cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant
20 according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes,
25 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or
30 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor).

The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12-49-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

(Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count
5 the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR
10 can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

15 NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of
20 apoptosis (NF- κ B appears to shield cells from apoptosis). B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target
25 genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in
30 treating diseases. For example, inhibitors of NF- κ B could be used to treat those

diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGAC
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
5':GCGGCAAGCTTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)
15 Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCC
20 ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGA
CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTA
TTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAA
GCTT:3' (SEQ ID NO:10)

25 Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP
30 cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly,

the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5

13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small
5 molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane
potential. These alterations can be measured in an assay to identify supernatants
which bind to receptors of a particular cell. Although the following protocol
describes an assay for calcium, this protocol can easily be modified to detect changes
in potassium, sodium, pH, membrane potential, or any other small molecule which is
10 detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to
measure changes in fluorescent molecules (Molecular Probes) that bind small
molecules. Clearly, any fluorescent molecule detecting a small molecule can be used
instead of the calcium fluorescent molecule, fluo-4 (Molecular Probes, Inc.;
15 catalog no. F-14202), used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star
black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for
20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of
HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

20 A stock solution of 1 mg/ml fluo-4 is made in 10% pluronic acid DMSO. To
load the cells with fluo-4, 50 ul of 12 ug/ml fluo-4 is added to each well. The plate
is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in
the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are
25 re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml
fluo-4 solution in 10% pluronic acid DMSO is added to each ml of cell suspension.
The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed
twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100
ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed
30 once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final
volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-4. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca^{++} concentration.

10

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

30

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr
5 with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell
10 number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

15 To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM
20 HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane
25 bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19; an assay which detects activation
5 (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other
10 phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G
15 plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at
20 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts
25 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place
of A431 extract. Plates are then treated with a commercial polyclonal (rabbit)
30 antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive

incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

5 **Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA
10 is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C. using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4
15 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

20 PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining
25 alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to
30 the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for

precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and

5 chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

10

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide

15 is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with

20 specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample

25 containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature.

30 The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day , and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 $\mu\text{g/kg/hour}$, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to

modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's

solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in

the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other
5 therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by
10 administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

15 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

20 **Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

25 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

30

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide. The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The

polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely

differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or
5 RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence
10 can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or
15 mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard
20 recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on
25 the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

30 After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A

time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

5 The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 28: Transgenic Animals.

10 The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are

15 used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (*i.e.*, polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear

20 microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40:691-698 (1994); Carver et al., Biotechnology (NY) 11:1263-1270 (1993); Wright et al., Biotechnology (NY) 9:830-834 (1991); and Hoppe et al., U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82:6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic

25 stem cells (Thompson et al., Cell 56:313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, *e.g.*, Ulmer et al., Science 259:1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-

30 mediated gene transfer (Lavitrano et al., Cell 57:717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115:171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380:64-66 (1996); Wilmut et al., Nature 385:810-813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89:6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265:103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse

transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 29: Knock-Out Animals.

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (*E.g.*, see Smithies et al., Nature 317:230-234 (1985); Thomas & Capecchi, Cell 51:503-512 (1987); Thompson et al., Cell 5:313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in

the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (*e.g.*, see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

10 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (*i.e.*, animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc. The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U.S. Patent No. 5,399,349; and

Mulligan & Wilson. U.S. Patent No. 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

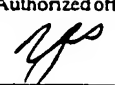
The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference. Further, the hard copy of the sequence listing submitted herewith and the corresponding computer readable form are both incorporated herein by reference in their entireties.

Applicant's or agent's file reference number	P2031PCT	International application No.	Unassigned
--	----------	-------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>259</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit July 27, 1998	Accession Number 203070
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

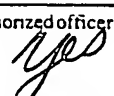
<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application	<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:
Authorized officer  Yvette E. Simms PCT International Division	Authorized officer

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
--	----------	-------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>260</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>July 27, 1998</u>	Accession Number <u>203069</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit") 	

For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer 	<u>Yvette E. Simms</u> PCT International Division


For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
--	----------	-------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>253</u> . line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>July 27, 1998</u>	Accession Number <u>203071</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit") 	

For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer 	Yvette E. Simms PCT International Division

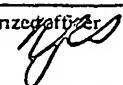
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>249</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit June 26, 1998	Accession Number 203027
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

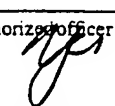
<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application	<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:
Authorized officer  Yvette E. Simms PCT International Division	Authorized officer

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>243</u> . line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit June 11, 1998	Accession Number 209965
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application
Authorized officer  Yvette E. Simms PCT International Division

<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:
Authorized officer

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

- (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
- (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
- (g) a variant of SEQ ID NO:Y;
- (h) an allelic variant of SEQ ID NO:Y; or
- (i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

- (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and
- (b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

<110> Human Genome Sciences, Inc.

<120> 98 Human Secreted Proteins

<130> PZ031.PCT

<140> Unassigned

<141> 1999-07-28

<150> 60/094,657

<151> 1998-07-30

<150> 60/095,486

<151> 1998-08-05

<150> 60/095,455

<151> 1998-08-06

<150> 60/095,454

<151> 1998-08-06

<150> 60/096,319

<151> 1998-08-12

<160> 364

<170> PatentIn Ver. 2.0

<210> 1

<211> 733

<212> DNA

<213> Homo sapiens

<400> 1

gggatccgga	gccc aaatct	tctgacaaaa	ctcacacatg	cccaccgtgc	ccagcacctg	60
aattcgaggg	tgcaccgtca	gtcttctctt	cccccccaaa	acccaaggac	accctcatga	120
tctcccggac	tcttgaggtc	acatgcgtgg	tgggtggacgt	aagccacgaa	gacctgagg	180
tcaagttcaa	ctggtacgtg	gacggcgtgg	aggtgcataa	tgccaagaca	aagccgcggg	240
aggagcagta	caacagcacg	taccgtgtgg	tcagcgtcct	caccgtcctg	caccaggact	300
ggctgaatgg	caaggagtac	aagtgcaagg	tctccaacaa	agccctccca	acccccatcg	360
agaaaaccat	ctccaaagcc	aaagggcagc	cccagagaacc	acaggtgtac	accctgcccc	420
catcccggga	tgagctgacc	aagaaccagg	tcagcctgac	ctgcctggtc	aaaggcttct	480
atccaagcga	catcgccgtg	gagtgggaga	gcaatgggca	gccggagAAC	aactacaaga	540
ccacgcctcc	cgtgctggac	tccgacggct	ccttcttctt	ctacagcaag	ctcaccgtgg	600
acaagagcag	gtggcagcag	gggaacgtct	tctcatgctc	cgtgatgcat	gaggctctgc	660
acaaccacta	cacgcagaag	agcctctctc	tgtctccggg	taaatgagtg	cgacggccgc	720
gactctagag	gat					733

<210> 2

<211> 5

<212> PRT

<213> Homo sapiens

<220>

<221> Site

<222> (3)

<223> Xaa equals any of the twenty naturally occurring L-amino acids

<400> 2

Trp Ser Xaa Trp Ser
1 5

<210> 3

<211> 86

<212> DNA

<213> Homo sapiens

<400> 3

gcgcctcgag atttccccga aatctagatt tccccgaaat gatttccccg aaatgatttc 60
cccgaaatat ctgccatctc aattag 86

<210> 4

<211> 27

<212> DNA

<213> Homo sapiens

<400> 4

gcggcaagct ttttgcaaag cctaggc 27

<210> 5

<211> 271

<212> DNA

<213> Homo sapiens

<400> 5

ctcgagattt ccccgaaatc tagatttccc cgaaatgatt tccccgaaat gatttccccg 60
aaatatctgc catctcaatt agtcagcaac catagtcccc cccctaactc cgcccatecc 120
gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa ttttttttat 180
ttatgcagag gccgaggccg cctcggcctc tgagctattc cagaagtagt gaggaggctt 240
ttttggaggc ctaggctttt gcaaaaagct t 271

<210> 6

<211> 32

<212> DNA

<213> Homo sapiens

<400> 6

gcgcctcgagg gatgacagcg atagaacccc gg 32

<210> 7

<211> 31

<212> DNA

<213> Homo sapiens

<400> 7

gcgaagcttc gcgactcccc ggatccgcct c 31

<210> 8

<211> 12

<212> DNA

<213> Homo sapiens

<400> 8

ggggactttc cc

12

<210> 9

<211> 73

<212> DNA

<213> Homo sapiens

<400> 9

gcggcctcga ggggactttc ccggggactt tccggggact ttccgggact ttccatcctg
ccatctcaat tag

60

73

<210> 10

<211> 256

<212> DNA

<213> Homo sapiens

<400> 10

ctcgagggga ctttcccgga gactttccgg ggactttccg ggactttcca tctgccatct 60
caattagtca gcaaccatag tcccggccct aactccggcc atcccggccc taactccggc 120
cagttccggc cattctccgc cccatggctg actaattttt tttatttatg cagaggccga 180
ggcgcctcgc gcctctgagc tattccagaa gtatgagga ggcttttttg gaggcctagg 240
ctttcgcaaa aagctt 256

<210> 11

<211> 1564

<212> DNA

<213> Homo sapiens

<400> 11

gcggacgggtg ggctgtgcaa ccttcctccc tttcttaaas gcttggggca tttgtctggc 60
cttccctttt actgtctggct gctgcctgca tctgtctctt aaccttcatt aactgtgcct 120
atgtcaaatg ggaaccctg gtacaagata ttttcacctt tgcataagta ttggcactga 180
tcgcgggtcat cgttgcaggc attgttagac ttggccaggg agcctctact ctttttgaga 240
attcctttga gggttcatca tttgcagtgg gtgacattgc cctggcactg tactcagctc 300
tgttctccta ctcaggctgg gacacctca actatgtcac tgaagagac aagaatcctg 360
agaggaacct gcccctctcc attggcatct ccatgcccac tgcaccatc atctatatct 420
tgaccaatgt ggctattat actgtgctag acatgagaga catcttggcc agtgatgctg 480
ttgtctgtgac ttttgcagat cagatatttg gaatatttaa ctggataatt ccactgtcag 540
ttgcattatc ctgttttggg ggcttcaatg cctccattgt ggctgtctct aggettttct 600
ttgtgggtc aagagaaggc catctccctg atgccatctg catgatccat gttgagcggc 660
tcacaccagt gccttctctg ctttcaatg gtatcatggc attgatctac ttgtgcgtgg 720
aagacatctt ccagctcatt aactactaca gcttcagcta ctggttcttt gtggggcttt 780
ctattgtggg tcagctttat ctgcgctgga aggagcctga tcgacctgt cccctcaagc 840
tcagcgtttt cttcccgatt gtcttctgcc tctgcaccat cttcctgggtg gctgttccac 900
tttacagtga tactatcaac tccctcatcg gcattgccat tgcctctca ggcttgcct 960
tttacttctt catcatcaga ttgccagaac ataagcgacc gctttacctc cgaagatcgt 1020
ggggtctgcc acaaggatcc tccaggctct gtgtatgtca gttgctgcag aaatggattt 1080
ggaagatgga ggagagatgc ccaagcaacg ggatcccaag tctaactaaa caccatctgg 1140
aatcctgatg tggaaagcag ggggtttctgg tctactggct agagctaagg aagttgaaaa 1200
ggaaagctca cttcttttga ggcacctgtc cagaagcctg gcctaggcag cttcaacctt 1260
tgaacttact ttttgaaatg aaaagtaatt tatttgtttt gctacatact gttccagact 1320
tttaaggggg acaatgaagg tgactgtggg gaggagcatg tcagggttgg gcttgggtgt 1380
tttagaagca cctgggtgtg cctacctact cctcttttct tttaaaaggg ccacaaatgc 1440

tccaattttcc	tgtctctctt	agagagacat	gaaactatca	caggtgctgg	atgacaataa	1500
aagtttatgt	tcctaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaagggcggc	1560
cccg						1564

<210> 12

<211> 1757

<212> DNA

<213> Homo sapiens

<400> 12

ggcacgaggt	agatggggtt	ttgggtgtgga	tgtcctttct	gtttgttagt	tgtcctttcta	60
acagacagga	ccctcagctg	caggtctgtt	ggagtacctt	gcaatgtgag	gtgtcagtgt	120
gccccgtctg	gaggggtgct	cccagttagg	ctgctcgagg	gtcaggggtc	agggaccac	180
ttgaggaggc	agtctgcccc	ttctcagatc	tcagctgca	tgtctggaga	accactgtct	240
tcctcaaagc	tgtcagacag	ggacatttaa	gtctgcagag	gttactgtct	tcctttttgt	300
tgtctgtgcc	ctgccccag	aggtggagcc	tacagaggca	ggcaggcctc	cttgagctgt	360
ggtgggctcc	acccagttcg	agcttccccg	ctgctttgtt	tacctaaaga	agcctgggca	420
atggcgggcg	ccccctcccc	agtctcgctg	ccgccttgca	gtttgatctc	agactgtctg	480
gctagcaatc	agcgagactc	cgtggggtag	gacctctga	gccaggtggg	ggatataatc	540
ttgtgtgtcg	ccatttttta	agcccgctcg	aaaagcgcag	tattcagggtg	ggagtgaacc	600
gattttcccg	attttccagg	tgcctgctgt	catccctttc	tttgattagg	aaagggaact	660
ccctgacccc	ttgcgcttcc	cgagtggagc	aatgccttgc	gctgctttgg	ctcgtgcatg	720
gtgcgtgcac	ccactgacct	gcgcccactg	tctggcactc	cctagtgaag	tgaacccggt	780
acctcagatg	gaaatgcaga	aatcacccgt	cttctgcgtt	gctcaggctg	ggagctgtag	840
accagagctg	ttcctattcc	gccatcttgg	ctcctcccc	accccccgcc	caatcctaag	900
tttatctctt	aacaaaagta	atactgcaca	cgtattttaca	tgttttaaaa	tattgtttat	960
ggagcactta	ctatgtgttt	tacgggtcaa	gtactttata	ttattcactg	actctttata	1020
gcaactctat	aacatattaa	tagattgtct	caatttcact	gatgacaaaa	ctggcagctt	1080
agtaactaat	gagtaacaat	accagcccct	aaaacctagg	gtctttccac	tacagcacac	1140
tctgactgac	attgctcttg	ttcagacaca	gcactcttca	cctagaaaag	gatggagagc	1200
tagaaaagtg	ttttcttttt	caagtgtgtc	ccattcatca	atattctatc	aagatatatg	1260
gaagagttaa	tttctaggtc	ttccagaagg	ggctgctgac	acttgggaat	tcaaaaataat	1320
tactccagag	ggaaatgtgg	gcatagaaac	tttttaaaaa	ctgaattttc	ccgatctaata	1380
atcgggttcc	acatacagta	tattattttc	taaaaatcag	ccagaaatgc	acctgcagtt	1440
ttatctactg	gatttttctt	tcccagcccc	attcttcttt	aagtagtttt	ttttcccat	1500
tttataaaaag	aaaagcatat	ctcccaccta	tagcttccct	ccttatggta	tattatctca	1560
ggggataatt	tgaaatattg	cttgagacac	aattctcatt	aagaaataaa	ctttcaggct	1620
gggtgcagtg	gctcacactt	gtaatcccag	aactttggga	ggccaaggcg	ggggagagtc	1680
cctgaaggcc	aagagtccaa	gaccagcctg	gaatatatac	caagatcccc	tctccacaaa	1740
aaaaaaaaaa	aaaaaaa					1757

<210> 13

<211> 1373

<212> DNA

<213> Homo sapiens

<400> 13

gtgaaatgaa	atatgtatac	caaaacacat	aaatttaagt	tttataattt	tccttccttg	60
tggaatctgga	aaattttctt	cttgctgttt	ttcattctga	tagttgcttt	ggctttccct	120
attccctgcc	tttccatatt	ttgaaatgtt	ttgggtcaaat	tttttcctgt	tgagtgtgca	180
gtatgttttg	attgttaatt	ttattgacca	tcattctcagt	ggattattta	tttgaaactt	240
ttctgtttcc	tggtctggaa	tggttcctgc	cttggaagcc	atattttaac	tgtatactga	300
tccagcttta	tttatcttaa	gtatttttac	aatttattta	ctgataatat	ataactaggt	360
ataatcactc	tagtcagtga	aatcaaacctg	ataacacaaa	atgatgctgc	tagtaaat	420
agcctgaaaa	ggatcagagg	aagcactttt	tagtctaagt	agtcaagatc	aataaactta	480
gaatgagagt	tttgtattat	ttagccttgt	gatagaagga	aaaagtacag	ttattgaatc	540
tgtatattaa	tttatgcaaa	atacacacat	tatctttcca	ttaacacagt	agtatgacat	600

acagttttat	caccactgga	ttcatgtatg	cttaaaatga	aaatgaaagg	aacacacagg	660
aaagtagctt	cttatcttcc	cagatacagc	ctccaagagt	gcctctgctg	tgtactggag	720
tgagtttgaa	gaagttcagg	taatcctctc	ctaaccagca	gcttaactta	aagtgcattc	780
tcatttttgt	cctaccacta	aaagcttgct	tattgacctc	cttctttcct	gtctcctaaa	840
ttgtgactgg	ctctcagtta	gacttgcatc	tggtgctatt	tgtagcagca	attgcaggat	900
tcatttgctt	ttctctctgc	agtgtatttc	tgaatgtttt	ataactctgg	gtatgcataa	960
tgtacgtata	tgcatatcca	tatggccaag	agttttcaga	attctttgtt	tgaataatga	1020
atgtgtgtat	tcctaccccc	ttgttttgc	ctgtcttaac	atttaagctt	tgcttattaa	1080
aatgtatttg	ttttaatgct	ttctctttag	attgtaagct	cttcatgggc	agtggtcaga	1140
tattctgtat	ttgtaacttc	ccgagaatct	agtattttat	catgtacata	atgaacatag	1200
aatagtattg	attggaagga	cagtgaagta	caaaaatgaga	ttctaaaatg	acattttcat	1260
tagggctcac	aagtgtataa	gaagtatccc	tttctatttt	ttttgtttaa	tactttagaa	1320
gctattacat	taaaaaagct	attctcatta	aaaaaaaaaa	aaaaaaaaaa	aaa	1373

<210> 14

<211> 3740

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (957)

<223> n equals a,t,g, or c

<400> 14

actctctaag	gcttttttgg	actcaccaaa	caggcttctt	gcagtggaga	tgaatacaga	660
tcacttaagg	ttgactgtgc	caaattggcat	aggggccctg	aagctaaggg	aaatggaaca	720
ctactttctc	cagggcctgt	cagttcagct	gtttaatgat	gggtccaagg	gcaaactcaa	780
tcattttatg	ggagctgact	ttgtgaaaag	tcatacagaaa	cctccacagg	gaatggaaat	840
taagtccaat	gaaagatgct	gttcttttga	tggagatgca	gacagaattg	tttattacta	900
ccatgatgca	gatggccact	ttcatctcat	agatggagac	aagatagcaa	cgtaatttag	960
cagtttctct	aaagagctcc	tgggtggagat	tggagaaaag	ttgaatattg	gtgttgatca	1020
aactgcatat	gcaaattgga	gttcaacacg	gtatcttgaa	gaagttatga	aggtacctgt	1080
ctattgcact	aagactgggt	taaaacattt	gcaccacaag	gctcaagagt	ttgacattgg	1140
agttttattt	gaagcaaatg	ggcatggcac	tgcactgttt	agtacagctg	ttgaaatgaa	1200
gataaaaaca	tcagcagaac	aactggaaga	taagaaaaga	aaagctgcta	agatgcttga	1260
aaacattatt	gactttgtta	accaggcagc	tgggtgatgc	atttctgaca	tgctgggtgat	1320
tgaagcaatc	ttggctctga	agggcttgac	tgtacaacag	tgggatgctc	tcatacaga	1380
tttccaaac	agacaactta	aagttcaggt	tgcagacagg	agagttatta	gcactaccra	1440
tgctgaaaga	caagcagtta	caccccaggg	attacaggag	gcaatcaatg	acctgggtgaa	1500
gaagtacaag	ctttctcgag	cttttgtccg	gccctctggt	acagaagatg	tcgtccngag	1560
tatatgcaga	agcagactca	caagaaagtg	cagatcacct	tgcacatgaa	gtgagcttgg	1620
cagtatttca	gctggctgga	ggaattggag	aaaggcccca	accaggtttc	tgaagataat	1680
tttcataatt	ctgagaaact	ggacttttta	caagtcttta	caaaactgtc	aataataatg	1740
gcagtactaa	gagatttata	atcataatgt	ttacaatgca	gctactgga	ttgtctctag	1800
atctgttttt	cttaaacact	aacagaataa	ttctttataa	ataggttaagc	cttacacttg	1860
ttaaagaaat	ttacctctaa	tttcagtcct	actaatgtaa	aatactggga	cttaagtata	1920
caattcagtc	actaaactgt	cagttttatg	tggggaacaa	ttcatgcagg	ctactggaaa	1980
attaaatctt	attaccaact	ccttctgata	tttttgccat	caccatcaca	tgagcaagat	
gatgttttgc	agcattcccc	attgctgata	caaatggaga	gggcagagaa	gacttttatac	
aaccagtttt	tccattgcag	agtcttaaga	aagattatta	gatgacttac	ctatatgact	
aatgccatca	ggaactcaga	ggtatgaata	gggggttgtc	catccctctt	ccatactgag	
gtggagatgc	tcattgcaata	cttttaagga	tgcattggct	agccttcagt	tattcttcac	
tgctcttggt	gaaggtatgt	gggagaaaaa	ctaattataa	tacgtttccc	agcctctgat	
ggagaaggaa	caccattctg	ataccagaac	atggttaata	aggaaaagag	aaaaatcccc	
aaccaatctt	aattgaacca	agtctgaaac	caatggaaaa	aaaaaatggg	tagtgtatat	
tttgagggtt	taagacaact	caggacaata	aaaacaatgg	actttacatg	tgtatatata	
tagctctctt	aggcaccata	atcagtatga	gccacaataa	tttaaacctg	attcaggcca	

cattcagaca	tttgctctta	tatacaaaata	tttaaatttaa	atacaatctg	aaatgtgttc	2040
tggtacatac	aaaaaaggaa	aaactataca	acgcagagca	gtgtgtgtgt	tttaaataat	2100
tacattttaca	tgtaagctaa	atggaaccag	caatgggtgt	caagttttta	tcattcccttc	2160
cagaaaatct	ttttctacca	tctcttctat	tttttgcttg	gctttgctgg	aacatgggttt	2220
gtggtttctc	agtttcatgt	ccttatttagg	gaaggcattt	gagtagagga	taggactccc	2280
tgagtgtcct	ccacatcggc	ttgtgacttt	gctgttgaag	acttgactga	gcacattgaa	2340
gaacggcagg	agctgctcca	tactgcgcac	gctgcagatg	gtgagcagca	agtgccttgg	2400
ctcccaaccc	aatgttctcc	ctgagttgtc	ttctcttggg	tttttctgca	gaaaacaaaa	2460
agtgaactgg	tattaatata	acagacaatg	tggtatgtta	gaaaaattaa	aaatatataa	2520
actttggcaa	ttggtcaaga	aatgaatata	aatgacatta	agtttctaac	tcctgacctg	2580
atcaaaaccc	ttggtgcttc	tgagaccttt	tactgccatt	tattagtttt	acatggagca	2640
gtctaacttt	gtagtaatat	ttcccaacta	gaatgcgcag	ataagcttag	ttaacagaaa	2700
tagcttttga	caggaataga	gtcaaacata	aaagttttat	gttggtgctt	gtatttactc	2760
aaaaagctcc	caggtttctg	aaccctcact	actgtaacca	aggactaggt	cacaaaatta	2820
ctacagaaaa	aaggaacaaa	gtgctttcata	catttcataa	tatatccctt	tttattataa	2880
ttagttaatt	ccctttttat	taaatggcct	aaatttgcca	tgatggtagc	agtgtccaaa	2940
gtgaataatt	actgtcagta	ctgcatcaca	gagaaaggaa	gggatccctc	aggagacact	3000
gctgtctcct	tctgggttgt	gctaaacaac	ataggaggga	aagctggacc	tggagtcaaa	3060
ggaattgagt	tagtgtgtgt	gctctgccat	acttacggca	cccttgggca	ggatatataa	3120
aggttctcta	cttataaaaat	gggacagtct	aaaactacct	tttagtagag	aagtcaaatg	3180
agaagggtatg	tgaaaaactct	gtcaactaaa	tataaagact	aataatttgg	gtattaagag	3240
gctagtttga	gaagccacct	gaattacaca	aacacagcta	cagacatcat	tctgtctaga	3300
gaaagataag	agagaacagg	ttggttgaac	ttgggcagaa	tcacagatac	aattccacac	3360
taaagaatga	aaataagcaa	tgaactagac	agaagggaaga	aatcatgaag	acttaggaag	3420
cagaattaca	atctgtcata	ttaacaaatg	gagtttgctt	tctaagatca	gatgttgctc	3480
agaaactttc	attgtttacc	taataaattt	atatcactag	tttctagtgt	ggtcaagcag	3540
atgcaaaatc	cagcttattt	tcttctatgt	gctctcaagc	ttattgctta	ttttaaagta	3600
aaatcctgaa	aaaggaaaat	attagggttg	tgcaaacgta	attgctggtt	ttgcatttgt	3660
gaaatttgcc	gttttatatt	ggagtacatt	cttaataaaa	tgtgggttat	ttatacaaaa	3720
aaaaaaaaaa	aaaactcgag					3740

<210> 15

<211> 1196

<212> DNA

<213> Homo sapiens

<400> 15

ccacgcgtcc	ggtgaatctg	gatgatctat	ttatgtcatt	tattctatgtg	tctcttctta	60
tgccagttcc	ccactgtctt	gattattgtg	gctttatggt	gtcttgaaat	taagtattgt	120
taagctttgc	aatcttttcc	aagattgttt	tggttatagg	tcctttcttt	ataaagttta	180
tattaagctt	ctccattttt	ctttaagaaa	aaatctgctg	gaatggcact	gaatttttag	240
gtcgattttg	gaagaattga	caatatggaa	cctttcaatc	gatggacatg	gtatgtttct	300
gcattttatt	gggcctttta	aaatttatct	cagcaatatt	ttgtagtttt	agtgaagaag	360
tcttgtatat	atcttttgtt	aaatgtattc	ctaaatattt	tgtggaaaatg	ttactgtaaa	420
tagtacttta	atttcatttt	taagttratt	gctgggtatac	agaaatatgt	tttatttttg	480
taatatggac	ttttatatct	tgtaattctta	tttaaaattca	cttagttcag	tagtagttct	540
tattttttatt	ttatttgaag	tgctttaggg	ataccttggtg	tacacaatcc	atgtcatata	600
tgaataacca	acaggtttac	cttttgtttc	ttcttgcttt	gtcgtaatgg	ctaggacctt	660
ccaacctaaa	tgggtgaata	gaagcgggtg	gtgagtttat	tactcttgct	ttattctcaa	720
cagcagagga	tagcgtctag	tttttcaata	ttaaaaatag	ttagctatag	gttttttccg	780
tttcagtgtt	ttatcagatt	gaagatccct	tttatcccta	ttttgccaag	agtgtcatga	840
atgcatgttg	aatttcacca	agtacttttt	cctgcatctg	ttgagagaat	cttgactttt	900
ctctttttatt	ttgttaattg	agattcttta	atgttaaact	gatcttgcat	tcctgatata	960
aaccatactt	agtcattgata	tgttatcccc	tttatataat	gctggattcg	gtttgcaaat	1020
attttgtttg	atttttgcatt	ctatgtctcat	gagacagatt	ggctctgtgat	tttccctttct	1080
tgtaagtctt	gtcaggtgtc	agggtctggg	taatttttgac	acattttacaa	acatatattt	1140
ttaagataaa	aataaaacttc	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa	1196

<210> 16
 <211> 2209
 <212> DNA
 <213> Homo sapiens

<400> 16
 gagctgcgcc gggcttgggc gcctggggcc gccgctcccc accgtcgttt tccccaccga 60
 ggccgaggcg tcccgagtc atggccggcc tgaactgcgg ggtctctatc gacttgctag 120
 gggttctgct gctgggtgcg gcgcgcctgc cgcgcggggc agaagctttt gagattgctc 180
 tgccacgaga aagcaacatt acagtcttca taaagctggg gaccccgact ctgctggcaa 240
 aaccctgtta catcgtcatt tctaaaagac atataaccat gttgtccatc aagtctggag 300
 aaagaatagt ctttaccttt agctgccaga gtccctgaga tcactttgtc atagagatcc 360
 agaaaaatat tgactgtatg tcaggcccat gtcccttttg ggaggttcag cttcagccct 420
 cgacatcggt gttgcctacc ctcaacagaa ctttcatctg ggatgtcaaa gctcataaga 480
 gcacgggttt agagctgcag ttttccatcc ctgcctgag gcagatcggc cggggtgaga 540
 gctgcccgaga cggagtcact cactccatca gcggccgaat cgatgccacc gtggtcagga 600
 ttggaacctt ctgcagcaat ggcactgtgt cccggatcaa gatgcawgaa ggagtgaana 660
 tggccttaca cctcccatgg ttccacccca gaaatgtctc cggcttcagc attgcaaacc 720
 gctcatctat aaaacgtctg tgcattcatg agtctgtgtt tgagggtgaa ggctcagcaa 780
 cctgatgtc tgccaactac ccagaaggct tccctgagga tgagctcatg acgtggcagt 840
 ttgtcggttc tgcaacctg cgggccagcg tctccttctt caactcaac ctctccaact 900
 gtraggagaa ggaggagcgg gttgaatact acatcccggg ctccaccacc aaccctgagg 960
 tgttcaagct ggaggacaag cagcctggga acatggcggg gaactcaac ctctctctgc 1020
 aaggctgtga ccaagatgcc caaagtccag ggatccctcg gctgcagttc caagttttgg 1080
 tccaacatcc acaaatgaa agcaataaaa tctacgtggt tgacttgagt aatgagcgag 1140
 ccattgtcact caccatcgag ccacggcccg tcaaacagag ccgcaagtct gtccctggct 1200
 gtttcgtgtg tctagaatct cggacctgca gtagcaacct caccctgaca tctggctcca 1260
 aacacaaaat ctcttctctt tgtgatgatc tgacacgtct gtgatgaat gtggaaaaam 1320
 ccataagytg cacagaccac cgtactgccc aaaggaaatc ctactcacty caggtgcccc 1380
 gtgacatcct ycamctgcct gtggagctgc atgacttctc ctggaagctg ctggtgcccc 1440
 aggcacaggc cagcctgggt ctggtgccag ccagaaagct gcagcagcat acacacgaga 1500
 agccctgcaa caccagcttc agctacctcg tggccagtg ccatccccagc caggacctgt 1560
 acttcggctc cttctgcccg ggaggctcta tcaagcagat ccagggtgag cagaacatct 1620
 cggtgacctt tcgcaccttt gccccagct tccgacaaga ggctccagg cagggtctga 1680
 cgggtgtcct tataccttat ttcaaagagg aaggcggttt caggtgacc cctgacacaa 1740
 aaagcaaggt ctacctgagg acccccaact gggaccgggg cctgccatcc ctacacctg 1800
 tgtcctggaa catcagcgtg ccagagagac aggtggcctg cctgactttt ttaaggagc 1860
 ggagcggcgt ggtctgccag acagggcgcg cattcatgat catccaggag cagcggacce 1920
 gggctgagga gatcttcagc ctggacgagg atgtgctccc caagccaagc tccaccatc 1980
 acagcttctg ggtcaacatc tytaaytgma gccccacgag cggcaagcag cttagacctg 2040
 tcttctcggg gacacttacc ccaaggactg tggacttgac tgtcatcctc atcgacggg 2100
 tgggaggtgg agtcttactg ctgtctgccc tcgggctcat catttgctgt gtgaaaaaaa 2160
 aaaaaarama aacaaggggc cccgctgtgg gtatctacaa tggcaacat 2209

<210> 17
 <211> 1774
 <212> DNA
 <213> Homo sapiens

<400> 17
 taggcacccc aggtcttttac actttatggt ccgggttcgta tgttgtgtga attgtaaccg 60
 gataccaatt tcacacaagg aaccagctta tgaccatgat tacgccaagc tcgaaattaa 120
 ccctcactaa agggaaacaaa agctggagct ccaccgcggt ggcggccgct ctagaactag 180
 tggatcccc gggctgcagg aattcggcac gagcggctgc ggcgcgagg tgagggggcg 240
 gaggttccca gcaggatgcc ccggctctgc aggaagctga agtgagaggc ccggagaggg 300
 cccagccgc cggggcagg atgaccaagg cccggctgtt cgggtgtgg ctggtgctgg 360
 ggtcgggtgt catgatcctg ctgatcatcg tgtactggga cagcgcaggc gccgcgcact 420

tctacttgca	cacgtccttc	tctaggccgc	acacggggcc	gccgctgccc	acgcccgggc	480
cggacaggga	cagggagctc	acggccgact	ccgatgtcga	cgagtttctg	gacaagtttc	540
tcagtgtctg	cgtgaagcag	agtgaacctc	ccagaaagga	gacggagcag	ccgcctgcgc	600
cggggagcat	ggaggagaac	gtgagaggct	acgactggtc	cccgcgcgac	gcccggcgca	660
gcccagacca	gggcccggcag	caggcgggagc	ggaggagcgt	gctgcggggc	ttctgcgcga	720
actccagcct	ggccttcccc	accaaggagc	gcgcattcga	cgacatcccc	aactcggagc	780
tgagccacct	gategtgga	gaccggcacg	gggccatcta	ctgctacgtg	cccaagggtg	840
cctgcaccaa	ctggaagcgc	gtgatgatcg	tgctgagcgg	aagcctgctg	caccgcggtg	900
cgccttaccg	cgacctcgctg	cgcattcccc	gcgagcacgt	gcacaacgcc	agcgcgcacc	960
tgacctcaa	caagttctgg	cgccgctacg	ggaagctctc	ccgccacctc	atgaagggtca	1020
agctcaagaa	gtacaccaag	ttcctcttctg	tgccgcgaccc	cttcgtgcgc	ctgatctccg	1080
ccttccgcag	caagttctgag	ctggagaacg	aggagtctta	ccgcaagttc	gccgtgcccc	1140
tgctgcggct	gtacgccaac	cacaccagcc	tgccgcctc	ggcgcgcgag	gccttccgcg	1200
ctggcctcaa	gggtgccttc	gccaacttca	tccagtaact	gctggaccgc	cacacggaga	1260
agctggcgcc	cttcaacgag	caactggcgcc	agggtgtaccg	cctctgccac	ccgtgccaga	1320
tcgactacga	cttcgtgggg	aagctggaga	ctctggacga	ggacgcgcgc	cagctgctgc	1380
agctactcca	gggtggaccg	cagctccgct	ttcccccgag	ctaccggaac	aggaccgcca	1440
gcagctggga	ggaggactgg	ttcgccaaga	ttccccctggc	ctggaggcag	cagctgtata	1500
aactctacga	ggccgacttt	gttctcttctg	gtaccccaa	gcccgaatac	ctcctccgag	1560
actgaaagct	ttcgcgcttc	ttttctctgc	gtgcctggaa	cctgacgcac	gcgcactcca	1620
gtttttttat	gacctacgat	tttgcaatct	gggcttcttg	ttactccac	tgctctctac	1680
cattgagtac	tgtatcgata	ttgtttttta	agattaatat	atttcaggta	tttaatacga	1740
aaaaaaaaaa	aaaaaaactc	gagggggggy	ccgg			1774

<210> 18

<211> 1674

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (1649)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1663)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1665)

<223> n equals a,t,g, or c

<400> 18

caagttggta	cgctgcagg	taccgggtccg	gaattccccg	gtcgaccac	gcgtccgggtc	60
gaagataggt	tcgagacgg	ggatgttgca	gctgatcatg	cagttggggt	cggtgctgct	120
cacacgctgc	cccttttggg	gctgcttcag	ccagctcatg	ctgtacgctg	agagggctga	180
ggcacgccc	aagcccagca	tcccagtgc	ttacctgtat	ttcgacatgg	gggcagccgt	240
gctgtgcgct	agtttcatgt	cctttggcgt	gaagcggcgc	tggttcgcgc	tgggggccgc	300
actccaattg	gccattagca	cctacgccc	ctacatcggt	ggctacgtcc	actacgggga	360
ctggctgaag	gtccgtatgt	actcgcgcac	agttgccatc	atcgccgggt	ttcttgtgtt	420
ggccagcgg	gctggggagc	tgtaccgcgc	gaaacctcgc	agccgctccc	tgcagtccac	480
cggccagggt	ttcctgggta	tctacctcat	ctgtgtggcc	tactcactgc	agcacagcaa	540
ggaggaccgg	ctggcgtatc	tgaaccatct	cccaggagg	gagctgatga	tccagctgtt	600
cttcgtgctg	tatggcatcc	tgcccttggc	ctttctgtca	ggctactacg	tgaccctcgc	660
tgcccagatc	ctggctgtac	tgctgcccc	tgctcatgctg	ctcattgatg	gcaatgtttgc	720
ttactggcac	aacacgcggc	gtgttgagtt	ctggaaccag	atgaagctcc	ttggagagag	780

tgtgggcac	ttcggaaactg	ctgtcatcct	ggccactgat	ggctgagttt	tatggcaaga	840
ggctgagatg	ggcacaggga	gccactgagg	gtcaccctgc	cttcctcctt	gctggcccag	900
ctgctgttta	tttatgcttt	ttggctctgt	tgtttgatct	tttgcttttt	taaaattggt	960
ttttgcagtt	aagaggcagc	tcattttgtcc	aaattttctgg	gctcagcgct	tgggagggca	1020
ggagccctgg	cactaatgct	gtacagggtt	ttttcctgtt	aggagagctg	aggccagctg	1080
cccactgagt	ctcctgtccc	tgagaaggga	gtatggcagg	gctgggatgc	ggctactgag	1140
agtgggagag	tgggagacag	aggaagggaag	atggagattg	gaagtgagca	aatgtgaaaa	1200
attcctcttt	gaacctggca	gatgcagcta	ggctctgcag	tgtgttttgg	agactgtgag	1260
agggagtgtg	tgtgttgaca	catgtggatc	aggcccagga	agggcacagg	ggctgagcac	1320
tacagaagtc	acatgggttc	tcagggtatg	ccaggggcag	aaacagtacc	ggctctctgt	1380
cactcacctt	gagagtagag	cagaccctgt	tctgctctgg	gctgtgaagg	ggtggagcag	1440
gcagtggcca	gctttgcccc	tctgtctgtc	tctgtttcta	gctccatggt	tggcctggtg	1500
ggggtggagt	tccctcccaa	acaccagacc	acacagtcct	ccaaaaataa	acattttata	1560
tagamaaaaa	aaaaaaaaaa	aaggggcgcc	gctctagagg	atccctcgag	gggcccagc	1620
ttacgcgtgc	atgcgacgtc	atagctctnt	ccctatagaa	gtngnaaagg	gttc	1674

<210> 19

<211> 2018

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (2010)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (2012)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (2014)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (2016)

<223> n equals a,t,g, or c

<400> 19

cagcaatgaa	atcctgcttt	cttttctcca	gaactactat	attcagtggc	taaatggctc	60
cctgattcat	ggttttgtga	atcttgcttc	ctttttttcc	aacctttgyt	tattttgtatt	120
gatgcccttt	gcctttttct	ttctggaatc	agaaggcttt	gctggcctga	aaaagggaat	180
ccgagccgcg	atttttagaga	cttttggtcat	gctttctctt	cttgcggtac	tcattcttgg	240
gatagtgtgg	gtagcttcag	cactcattga	caacgatgcc	gcaagcatgg	aatctttata	300
tgatctctgg	gagttctatc	taccttattt	atattcctgt	atatcattga	tgggatgttt	360
gttactttct	ttgtgtacac	cagtggcctt	ttctcgtatg	ttcacagtga	tgggtcactt	420
gctagtgaag	ccaacaattc	ttgaagacct	ggatgaacaa	atttatatca	ttaccttaga	480
ggaagaaagca	ctccagagac	gactaaatgg	gctgtcttca	tcggtggaat	acaacataat	540
ggagtgtgaa	caagaacttg	aaaatgtaaa	gactcttaag	acaaaattag	agaggcgaaa	600
aaaggcttca	gcatgggaaa	gaaatttggc	gtatcccgct	gttatgggtc	tccttcttat	660
tgagacatcc	atctcggtec	tcttgggtgg	ttgtaatat	ctttgcctat	tggttgatga	720
aacagcaatg	ccaaaaggaa	caaggggggc	tggaaatagga	aatgcctctc	tttctacgtt	780
tggttttgtg	ggagctgcgc	ttgaaatcat	tttgattttc	tatcttatgg	tgtcctctgt	840
tgtcggcttc	tatagccttc	gatttttttg	aaactttact	ccaagaaaag	atgacacaac	900
tatgacaaaag	atcattggaa	attgtgtgtc	catcttgggt	ttgagctctg	ctctgcctgt	960

gatgtcgaga	acactgggaa	tcactagatt	tgatctactt	ggcgactttg	gaagggttaa	1020
ttggctggga	aatttctata	ttgtattatc	ctacaatttg	ctttttgcta	ttgtgacaac	1080
attgtgtctg	gtccgaaaat	tcacctctgc	agttcgagaa	gaacttttca	aggccctagg	1140
gcttcataaa	cttcacttac	caaatacttc	aagggattca	gaaacagcca	agccttctgt	1200
aaatgggcat	cagaaagcac	tgtgagacgc	acagacggcg	tcttctgcca	ccaagagacc	1260
cgagaactcc	agattcacga	cattcctgtc	ccatgtagaa	gcatttccat	tcaaccgtgg	1320
ccccctttca	gaacctagac	ctatcagtgc	catttttttt	tcataatcta	cgaagaactt	1380
ggctatggct	gatctttttt	aaatttaact	ttctgatgga	ccctgtagtt	tccagttaag	1440
tgcagattcc	ttacagacat	atagaacagc	gcattcttct	gtagacattt	gctcatgttg	1500
gtaaatacaa	tcacccatat	gaaaaaattg	ttttcacctg	atatgaaaat	gttagaaaaa	1560
gcaaacctcc	ggactttctaa	agatttactt	aaatcccatt	atgtacttta	ttcagaatgt	1620
agaagctgac	ttgaaaggca	tccttggtag	taagtgaagc	ttattcagaa	aatgcatttt	1680
tcaaatgcaa	tggcaactgc	ttgtagatat	catttttgca	gtgtatgttg	gagctgtaat	1740
ggttgcaatt	atgtttctta	tttctttaa	agcaaaaagc	gtagtctctg	atttatgtta	1800
tagaatgata	ctgatttagac	tttgagccaa	ggggaaaata	ctaaattctt	ttaaaccctgg	1860
agccttagag	agccacagga	atatcttctg	ttgtacagtc	taataagctg	tggttaggaag	1920
tatcatgtaa	tcacagttta	atgacagttt	atgtatatat	ataattcagt	attccctcga	1980
ggggggggccc	ggtacccaat	tcgcccagn	tnancnag			2018

<210> 20

<211> 2098

<212> DNA

<213> Homo sapiens

<400> 20

ggcacgaggg	accgagctat	tctcctggga	ctggctatga	tgggtgtgctc	catcatgatg	60
tattttctgc	tgggaatcac	actcctgcgc	tcatacatgc	agagcgtgtg	gaccgaagag	120
tctcaatgca	ccttgctgaa	tgcgtccatc	acggaaacat	ttaattgctc	cttcagctgt	180
ggtccagact	gctggaaact	ttctcagtac	ccctgcctcc	aggtgtacgt	taacctgact	240
tcttcggggg	aaaagctcct	cctctaccac	acagaagaga	caataaaaaat	caatcagaag	300
tgctcctata	tacctaaatg	tggaaaaaat	tttgaagaat	ccatgtccct	ggtgaatgtt	360
gtcatggaaa	acttcaggaa	gtatcaacac	ttctcctgct	attctgacct	agaaggaaac	420
cagaagagtg	ttatcctaac	aaaactctac	agttccaacg	tgctgttcca	ttcactcttc	480
tggccaacct	gtatgatggc	tgggggtgtg	gcaattgttg	ccatgggtgaa	acttacacag	540
tacctctccc	tactatgtga	gaggatccaa	cggatcaata	gataaatgca	aaaatggata	600
aaataatttt	tgttaaagct	caaatactgt	tttctttcat	ttttcaccaa	agaaccttaa	660
gtttgtaacg	tgcagctctg	tatgagttcc	ctaatatatt	cttatatgta	gagcaataat	720
gcaaaaagctg	ttctatatgc	aaacatgatg	tttttattat	tcaggagaat	aaataactgt	780
tttgtgttgg	tgggtggttt	tcataatctt	atttctgtac	tggaaactagt	actttcttct	840
ctcattccgc	caaaacaggg	ctcagttatt	catttgccaa	gcttcgtgga	ggaatgtagg	900
tgacatcaat	gtgataaagt	ctgtgttctg	agttgtcaga	tctcttgaag	acaatatatt	960
tcatcactta	ttgtttacta	aagctacagc	caaaaaatat	tttttttctt	attctaaact	1020
gagccctata	gcaagtgaag	ggaccagatt	tcttaattaa	aggaagttag	gtacttttct	1080
tgtatttttt	accatatcac	tgtaaaagaag	aggggaaacc	cagccagcta	ctttttttca	1140
tcacttttta	ttcataactt	cagattttgt	aaactaattt	ccaaaatata	agctgttttc	1200
attagccagt	tctataatat	cttctgtgta	tttatgtaga	aaatgaacac	acctcttttc	1260
catttaagac	cctgctactg	tgtgaagaga	tgatacttac	aaggagtgtc	attacctgtg	1320
agctgactga	atgttggtag	gtgtccatt	acaatccagg	aaagctctgt	ttactgatat	1380
ttgtgtggaa	atctttattt	cacttcaatt	taaccattag	atggtaaaaat	taagatgcta	1440
cttgttggta	aaaattgggt	gactggtttc	aatgggtaaa	tgtgtgtggg	caaatataatg	1500
tgttggaaata	ttgtctcttg	tgaatttctg	cttaagtcaa	tgaatgtgta	gtatctctct	1560
ctgacaagca	ttccctattg	ggatttttaa	gctatgtgca	cagaatatta	gtctcttcta	1620
catgttttat	ttttctattt	ataattccct	tttttgttgt	tatattttat	acacagaata	1680
gatctttttt	ctaacacata	tttgaactga	ataacagact	taaagaaaagc	ctttgttcac	1740
attgtctattt	acttttgtgt	ttgggggaaa	atacagaggga	ttgattttta	ataaaaaaca	1800
ttccatcttt	catttaatat	caatatcaaa	agaagaagac	aaacatctat	ctttctcttc	1860
tatattttaag	tacttttttg	taatgtagta	tcaaagtttt	ttaggtaatg	caaaaatttta	1920
caaatcattt	gtggaatgaa	tggtaaaaact	aatctgatga	aatggaaaat	tattctgcaa	1980

<400> 21						
gcggttcgcgc	acctccagct	cgggccgatg	tggaagcttt	ggagagctga	agagggcgcg	60
gcggcgctcg	gcggcgcgct	cttcctgctg	ctcttcgcgc	taggggtccg	ccagctgctg	120
aagcagaggc	ggccgatggg	cttccccccg	gggccgcgcg	ggctgccatt	tatcggcaac	180
atctattccc	tggcagcctc	atccgagctt	ccccatgtct	acatgagaaa	gcagagccag	240
gtgtatcggg	aggtacagcc	ccgacggggc	ccggcgaggg	agggccgcga	agctggcccg	300
ggctggccag	ggccttcctg	cttggaactt	tggccgcccc	tggccgact	agtcgggacc	360
tctcctgtgt	ccggctgccc	tttgaggggac	acccgcttcc	cgggtctgga	agggagaagt	420
cctcgacgcc	gtgccccctt	gcaggggggag	ccccgcccc	gccggtgacc	cactccgggc	480
cgaggctccg	aggcgatcca	gtcctgattt	tcccgctacc	gctcgagctc	ttgctcctgc	540
gcttgcgccg	tttggtctcg	cagccgcgcc	gccacttcag	gtccagggtg	gacgcgatgc	600
ctcaggtcgc	ggcgctcttc	gagtcggcct	cgcagctctg	tggaaagctgc	acggcgcttg	660
tcgaaaaatc	aaggcgctct	gagttctaga	tgggtaatag	caggttctct	gtgtcttgca	720
gtcgacgaac	gactgggtga	ggcgcttctg	gtgagaatgg	agaatgcagg	ggaacgcccc	780
tgactgagaa	gcggggccctg	ggaaacgatt	gtgaacgcgt	gaatgaattg	atgactaaaa	840
tccgctgcgg	gggtcctaca	gcgcagatgg	taatgccgtt	ctgactgggt	gggaacggca	900
cctttagcaga	tacttaaaaag	gcgccttctg	tgtgccactg	tcaactgccaa	cttggtgact	960
catttaaaac	tataaaccag	ccggtgaggt	cggtaacttcg	ctctctctca	ctctgcggag	1020
gggaaagcag	cacggaaaatg	ccctgtgact	ggcagcggaa	aaggcgacca	cctgttgtgt	1080
gtgggtgtcc	cgacgtccgg	agggggcgagg	agttttccacg	ggtcctggga	cagagctcac	1140
ctgttttgtt	ttgaattaca	cttattttata	tgcaactaca	ggcctgacgc	tagcggtgaa	1200
gaaggcagat	acagcctttt	aaggagttag	cagatgagtg	ggagagagaa	aactaatctc	1260
attatcggcc	acaggctgtg	gtcagtgttt	tgaaggaaaa	gtacagggat	gtttggcaac	1320
tgtgttattt	caggtttgac	cttaaatcct	tacttaaac	agtttttaca	aggattgggtc	1380
taggtgcccg	ggcgcggtgc	tcacgcctat	aatctaacga	ctttgggaag	cggagcgggg	1440
cggatcacga	aatcaggaga	tcgagaccgt	cgtggctaac	acggtgaaac	cccatctcta	1500
ctaaaagaat	acaaaaaatt	ggccggggcg	ggtggcgggc	acctgtggtc	ccagctattc	1560
gggaggctgg	ggcaggagag	tggcgtgaac	ccgggagcg	gagctttcag	tgagccgaga	1620
tcggccact	gactccagc	ctgggcaaca	gagccagact	ccgtctcaaa	aaaaaaaaaa	1680
aaaaaggcg	gccgctctag	aggatccaag	cttacgtacg	cgtgcatgcg	acgtcaatag	1740
ctcttc						1740

<400>	22						
ccacgcgtcc	ggcggcgagg	cggcgggtggt	ggctgagtc	gtggtggcag	aggcgaaggc		60
gacagctcta	gggggttgga	cgggccccga	gaggaggatg	cgggtccgga	tagggctgac		120
gctgctgctg	tgtgcggtgc	tgctgagctt	ggcctcggcg	tccctcggatg	aagaaggcag		180
ccaggatgaa	tcccttagatt	ccaagactac	tttgacatca	gatgagtcag	taactgaccca		240
tactactgca	ggcagagtag	tgtctggcca	aatatttctt	gattcagaag	aaatggaatt		300
agaatcctct	attcaagaag	aggaagacag	cctcaagagc	caagaggggg	aaagtgtcac		360
agaagatatc	agcttttctag	agtctccaaa	tccagaaaac	aaggactatg	aagagccaaa		420
gaaagtacgg	aaaccaggta	gtctggacat	tttccttgct	ttttgattta	tttagggggac		480
aactgaaaaa	tttaagctaa	tgaataaaga	ggctgaagaa	gactggcttc	actgattatt		540
accacaaaat	aatatatgga	gtgtagtgtt	gagagaattt	ctagatttca	atataccaaa		600
gttattcact	taacagattt	gacccaagtt	acataagctg	aattccatag	atgagattgc		660

atttatcttt	ttattaatat	cctgacggtt	ctttttggaa	tcaagtatgt	aaataccttt	720
ctcactgaga	aattttaaaaa	aataatatgt	cagggtagca	tacaaaaacag	aaactcagga	780
agaactgtct	aacatgccac	tatgggttat	ttcaggattg	tggtaggatt	attactacaa	840
taaccttgtc	gatcatccta	ggtaatttga	tttaggtgtt	tatccctcag	ctgacgcttg	900
tgtgtacttc	catcctattc	taactttcac	aatcttgaaa	cctctacttt	tcaattttaa	960
acttgacctt	tttcacaata	gtctttcaaa	tattaatcct	tttaaaaaatt	attaatgtat	1020
gcattctacc	atgtcttggt	gcccttctgt	catttattat	ttcaatgatc	actgccttat	1080
ctgtttatta	ttccattact	aaccactctc	aggaactgat	ccaaatttga	acttttttaa	1140
aaaatagaat	ttttttttac	tacaaatata	ccattctcat	gaatagaaaa	ctatttttaa	1200
aaaattaaaa	gtacatgaat	cccatcaact	agaaataaca	gctgctcaca	tcattggagaa	1260
tattctctca	ggattttttt	aggtgtatca	gtagttttaa	gaaaagttaa	ttgggtcaaa	1320
atgtgtattc	tatgttgacg	cctgcatttt	gcacttaaca	gtttaccatg	aatgtttttc	1380
catgtcatct	agcttatttt	caatttgata	gttaattggc	ataaaaaatt	ttattttaa	1440
atatatggta	ccataaacca	aaacgtttat	gttttgctgg	gagatcattt	tagatgtatc	1500
tttgtgccta	tctgtgaaaa	tttctctaga	attcttaaa	taaatttgct	gaattggatg	1560
taaaagttta	aaagacagta	gatacacatt	gttaatttgt	tcttcagaaa	agtgattctg	1620
tttacttctt	ctgaaaatac	ctcttttccc	gtacctcacc	caagtgtgta	ttaccatttt	1680
aaaattttta	ctaaatataa	tttaattcat	tcttcttttt	ttaattaggg	attttttttt	1740
ttttttttta	gttttttatat	tttttagagac	atggtctcac	tctgtcactc	aggctggagt	1800
gcagtggcac	gatcatagtt	cactgcagcc	ttgaactcag	gtgatctccc	tctttggcct	1860
cccaaagtgc	tgggattaca	ggcatgagac	actacatcaa	gcctgggaaa	tttttaaa	1920
tgcataaaa	tagagggtaa	aatgacaccc	cttaccaacc	ctcactcagc	ttcaataaac	1980
atcaacatct	tgcctggctt	ttcatcttta	ctaccttaca	cacatttgcc	ctttttttct	2040
ttcctagaat	atttttaaggc	aaattccgca	tgtgctttta	catctctgtg	tctctctcta	2100
gtaagaatgt	tctttttcagc	aaaatcacag	aattagccat	tactccttta	ttaccttaata	2160
cccagactgt	ttaattttttc	tagttgcttc	aagtgtctgt	ttacaattta	tttgtttgaa	2220
ttcattttcta	agctactcag	gaggtcagc	caggagaatt	gcttgaaccc	aagaggcaga	2280
gggtgcagtg	agctgagatc	gtgccactgt	actccagcct	gggtgacaaa	gtgagactcc	2340
atgtcaaaa	aaaacaaaac	aaaacaaaac	aaaacaaaac	aaacaacaca	attgaataat	2400
atccctcagt	tattatattag	agaagagatg	actcaatata	aatgtaatat	gtgatcattt	2460
ggtcacagaa	tgaatttcag	tcattgtcta	ggtaagcctg	gtaaaaaaa	aaaaaaacgg	2520
gattccatga	ttagtgtgtt	gaaaaggggg	ttgccaaagt	gtaacagaca	gagccgggtg	2580
cagtggctca	cgctgtaat	cccagaactt	tgggaggcca	aggcaggcgg	atcacgaggt	2640
caagagattg	agacaatcct	ggccaacatg	gtgaagccct	gtatcaacta	aaaatacaaa	2700
aacaattagc	tgggcgtggt	gacatgtgcc	tgtagtccca	gctactcagg	aggctgaggg	2760
aggagaattg	cttgaacctg	ggaggcggag	gttgacgtga	gctgagattg	ctccactgca	2820
ctccagcctg	gccacagagt	gagattccat	ctcaaaaaaa	aaaaaaaaaa	aaaaaa	2876

<210> 23

<211> 1052

<212> DNA

<213> Homo sapiens

<400> 23

tcgaccacag	cgctccgcca	cgcgctccgc	cacgcgtccg	ggcggggcca	ggacgtgcac	60
tatggctcgg	ggctcgctgc	gccggttgct	gcggctccct	gtgctggggc	tctggctggc	120
gttgctgcgc	tccgtggccg	gggagcaagc	gccaggcacc	gccccctgct	cccgcggcag	180
ctcctggagc	gcgacctgg	acaagtgcac	ggactgcagc	acctcctgct	cccttccggc	240
tgctttggcc	catccttggg	ggcgctctga	gcctgacctt	cgctctgggg	ctgctttctg	300
gctttttggt	ctggagacga	tgccgcagag	agagaagttc	accaccccca	tagaggagac	360
cggcggagag	ggctgcccag	ctgtggcgct	gatccagtga	caatgtgccc	cctgcccagc	420
ggggctcgcc	cactcatcat	tcattcatcc	attcttagagc	cagtctctgc	ctcccagacg	480
cggcgggagc	caagctcctc	caaccacaag	gggggtgggg	ggcgggtgaat	cacctctgag	540
gcctgggccc	agggttcagg	ggaaccttcc	aagggtgtctg	gttgccctgc	ctctggctcc	600
agaacagaaa	gggagcctca	cgctggctca	cacaaaaacag	ctgacactga	ctaaggaaact	660
gcagcatttg	cacaggggag	gggggtgccct	cttctctaga	ggccctgggg	gccaggctga	720
cttggggggc	agacttgaca	ctaggcccca	ctcactcaga	tgtcctgaaa	ttccaccacg	780
ggggtcaccc	tgggggggtta	gggacctatt	tttaacacta	gggggctggc	ccactaggag	840

ggctggccct	aagatacaga	cccccccaac	tccccaaagc	ggggaggaga	tattttacttt	900
ggggagagtt	tggaggggag	ggtggggggg	gggaaaaaaa	aataaaaaaa	aaaattttta	960
atttttaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1020
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aa			1052

<210> 24

<211> 1541

<212> DNA

<213> Homo sapiens

<400> 24

ggcagcagct	tccctgatcc	tactctgtct	ctctttgttt	tggagaagtt	caatttacca	60
gctggatatg	tgggactagt	attcctgggt	atggcactgt	cctatgccat	ctcttcacca	120
ctatttggtc	tcctaagtga	taaaaggcca	cctctaagga	aatggcttct	ggtgtttggc	180
aacttaatca	cagccgggtg	ctacatgtct	ttagggcctg	tcccaatctt	gcataataaa	240
agtcagctct	ggctgctggt	gctgatatta	gttgtaagtg	gcctctctgc	tggaaatgagt	300
ataattccaa	ctttcccga	aattctcagt	tgtgcacatg	aaaatgggtt	tgaagaggga	360
ttaagtacat	tgggacttgt	atcaggctct	tttagtgcaa	tgtggccaat	tgggtgcttt	420
atgggacca	cgctgggtgg	atctctgtat	gagaaaaattg	gttttgaatg	ggcagcagct	480
atacaaggtc	tatgggctct	gataagtgga	ttagccatgg	gcttggttta	tctactggag	540
tattcaagga	gaaaaaggtc	taaatctcaa	aacatccctca	gcacagagga	ggaacgaact	600
actctcttgc	ctaataaaac	ctagtcggat	ggatccctga	ttgatacaag	gttgagaaat	660
gaatgctcct	ggccttaaac	atcaccgtag	gaagggtttt	taaaatttta	cgcgcaaaac	720
tccgtggacc	ccgtgccagt	gtcttgggaag	tgtcaacgtg	tttttggatg	atcctgtatt	780
gggctgtact	tactgtgata	ctgaaaagct	gtcctgctga	agcagctata	tttgaaatat	840
taagtatgaa	aggagtaatt	aaaaacaagc	aaaacaaaac	aagacttagt	ttttaaatga	900
ccaaacttgt	ctttaaagat	gttgttatta	actcgagtta	gttcttattt	cctctgttta	960
ttttttattc	taagtacact	gattctgtga	atgtaccttt	tttattacac	agggaaagaa	1020
atgaattaat	ttgatattgt	ctaaatacat	aaagggtgct	caaaatatgt	agaaacatta	1080
ctatgaaatc	agttttttaa	agataacttt	tctctttgtc	ctgagggttt	tcgggtctgt	1140
tcaaaaaggaa	gaattcttgc	ctgccataca	gaaactctct	agcaactccct	gaccttaagc	1200
ttttctaaaa	attctgtttg	tgtgaaaagt	acaagaataa	caatacttac	aacttccatt	1260
tttgtaacct	acgttcactt	atgatctgga	tttataaaca	ttacttggtg	taacgttttt	1320
catttccttt	aatgtctctg	ttttttggct	ctacctcttg	ttttgttttt	gttttttatct	1380
atatcttggg	agatgtattt	catccctaga	gcaggctcagc	ctctctcccc	taatgcgaat	1440
gcttgttttt	ttagggaagg	gcttctctca	acttctgttg	aaattgtgat	gttgaagtga	1500
ataaatgtct	attgtgtaac	aaaaaaaaaa	aaaaaaaaaa	a		1541

<210> 25

<211> 2079

<212> DNA

<213> Homo sapiens

<400> 25

ggcagcaggg	aggcgcggt	gcgggacctg	actagattct	acgacaaggt	actttctttg	60
catgaggatt	caacaacccc	tgtggctaac	cctctgcttg	catttactct	catcaaacgc	120
ctgcagctct	actggaggaa	tgtggtaaat	agtctggagg	ccagtggaga	catccgagct	180
ctgaaggatg	gctatgagaa	ggtggagcaa	gaccttccag	cctttgagga	ccttgaggga	240
gcagcaaggg	ccctgatgcg	gctgcaggac	gtgtacatgc	tcaatgtgaa	aggcctggcc	300
cgagggtgtct	ttcagagagt	cactggctct	gccatcactg	acctgtacag	ccccaaacgg	360
ctcttttctc	tcacagggga	tgactgcttc	caagttggca	agggtggccta	tgacatgggg	420
gattattacc	atgccattcc	atggctggag	gaggctgtca	gtctcttccg	aggatcttac	480
ggagagtggg	agacagagga	tgaggcaagt	ctagaagatg	ccttggatca	cttggccttt	540
gcttatttcc	gggcaggaaa	tgtttcgtgt	gcccctcagcc	tctctcggga	gtttcttctc	600
tacagcccag	ataataagag	gatggccagg	aatgtcttga	aatatgaaag	gctcttggca	660
gagagcccca	aacactgtgt	agctgaggct	gtcatccaga	ggcccaatat	acccaccttg	720
cagaccagag	acacctacga	ggggctatgt	cagaccctgg	gttcccagcc	cactctctac	780

cagatcccta	gcctctactg	ttcctatgag	accaattcca	acgcctacct	gctgctccag	840
cccatccgga	aggaggtcat	ccacctggag	ccctacattg	ctctctacca	tgacttcgtc	900
agtgactcag	aggctcagaa	aattagagaa	cttgacagaa	catggctaca	gaggtcagtg	960
gtggcatcag	gggagaagca	gttacaagt	gagtaccgca	tcagcaaaa	tgccctggctg	1020
aaggacactg	ttgacctaaa	actggtgacc	ctcaaccacc	gcattgctgc	cctcacaggc	1080
cttgatgtcc	ggcctcccta	tgacagat	ctgcagggtg	tgaactatgg	catcggagga	1140
cactatgagc	ctcactttga	ccatgctacg	tcaccaagca	gccccctcta	cagaatgaag	1200
tcaggaaacc	gagttgcaac	atttatgac	tatctgagct	cggtggaagc	tggaggagcc	1260
acagccttca	tctatgccaa	cctcagcgtg	cctgtggtta	ggaatgcagc	actgttttgg	1320
tggaaacctgc	acaggagtg	tgaaggggac	agtgcacac	ttcatgctgg	ctgtcctgtc	1380
ctggtgggag	ataagtgggt	ggccaacaag	tggatacatg	agtatggaca	ggaattccgc	1440
agaccctgca	gctccagccc	tgaagactga	actgttggca	gagagaagct	ggtggagtc	1500
tgtggctttc	cagagaagcc	aggagccaaa	agctggggtg	ggagaggaga	aagcagagca	1560
gcctcctgga	agaaggcctt	gtcagctttg	tctgtgcctc	gcaaatcaga	ggcaaggag	1620
aggttggttac	caggggacac	tgagaatgta	catttgatct	gccccagcca	cggaagtcag	1680
agtaggatgc	acagtacaaa	ggagggggga	gtggaggcct	gagagggaag	tttctggagt	1740
tcagatactc	tctgttggga	acaggacatc	tcaacagtct	caggttcgat	cagtggttct	1800
tttggcactt	tgaaccttga	ccacagggac	caagaagtgg	caatgaggac	acctgcagga	1860
ggggctagcc	tgactccagc	aactttaaga	ctttctcccc	actgccttct	gctgcagccc	1920
aagcagggag	tgtccccctc	ccagaagcat	atcccagatg	agtgtgtacat	tatataagga	1980
ttttttttta	gttgaaaaca	actttctttt	ctttttgtat	gatgggtttt	taacacagtc	2040
attaaaaatg	tttataaatc	aaaaaaaaaa	aaaaaaaaaa			2079

<210> 26

<211> 1947

<212> DNA

<213> Homo sapiens

<400> 26

tgtaaacaga	ttggagaatc	tagcaataag	attcaaagct	aatctggagc	ataaaggcac	60
agttcagaga	cagaataaca	gggatcacaa	gcatgaatta	aaaggaaattt	atttgccttca	120
agttcctaga	tacaaccttc	ccatgctgca	cttctccact	gtcggagcac	gttccgaaaa	180
acagaatgcc	ttgatccctg	gtgggtgcga	aggcagttgt	tagggatggc	aggcattgggt	240
gggctccaaa	agatgaaggc	cccacacaca	ggtgtgctgc	atttgggac	tgtgtgggtg	300
tttcttggac	cctttcttct	gggagtaggg	tacacactaa	cgtttaatcc	gctgtctggg	360
tgcattgcca	cagtagcggg	gctaaactcg	aacatcactg	caaataaggac	gctgagcagg	420
tccgtctgtc	atgtcacgcc	actgcacagg	tcttctgccc	cacacgacgg	ggagtaattg	480
cgtcagatgt	tattgaatag	ctcgtctcgg	gcaggggaa	cggggagttg	gggataataa	540
ttgggggttt	taattctatt	atcatgtcag	ctgacattat	gactataata	tgtagttaga	600
gacaattttt	atcttgctta	tagtaaaggt	tcagcctgcc	aattgtaaat	cattctaatt	660
tggcaggcct	atttttgaca	ttggaaagg	cagaaagcga	tttgcctccag	tagtgtaata	720
ggagttatag	accagaggct	gaaacccaaa	ctatataaaa	aggaatttcag	tggagggggc	780
tttgtaatct	ccattaatct	gtgttgctac	ttccaggatc	accaaaaatt	acatgtaatt	840
ttacatgtta	aacacattga	aacataacct	atgtttataa	agcataacgg	gcttcccttc	900
cagaagctct	cctgcttctc	atgaagttag	aacaatgaaa	agtcatagca	gatactcagt	960
ttaactctgt	gtagaacctc	gtagtgtttg	agctgttatt	cagatttgaa	ttcagactgt	1020
gtgttgtttg	cttatggaca	ctgcctgtcg	ttctgtcact	gttaaatata	tgagtctata	1080
agggtttttt	tccagaggcc	ataggtgaca	tcactaaaat	tgcaagataa	attgtaattc	1140
ttgctgtctg	tgcactcccc	aacctctccc	ccaccccccg	tggtgtgctg	ctttctagat	1200
gagcgtgttt	tggagcaggc	ccatctggga	cactctatgc	tttcaccaag	gaagtgcgat	1260
ctgagcagcc	acaatccagc	caaaagagga	tcgtagatat	ttgctctgat	caactagatg	1320
aaaatatagc	agaatggatt	tagccactcg	ctctgtttta	tccaactgag	tctctgacca	1380
gcaattgggtg	cataattatt	acagcaaaa	taaagaaatg	aaactgtagc	aattatgtaa	1440
atgaatgtgt	tggcctctta	atacctgtta	ctagtggact	tcctgtgagg	aagttagttt	1500
tttgttttga	tgaatgctt	tcgtttttta	aatcttaatt	ctgctgtcca	cattctccca	1560
aagtgtgctt	acttcatttg	tttaatttaa	atgaactttc	ctccttgtat	gtatgaggtg	1620
acttggtggg	tgggggtggg	gggttttgtt	tttgtgtttt	ttctttctta	gggcatctgt	1680
aggcctcaaa	ggacctttcc	tttaggtcat	attcctcaga	aagtcctcaa	tcttcccttg	1740

tttttgtttg	tttgtttttc	ttaaagaata	ttttcaaagc	ttaaatttgt	atattaattt	1800
aggactat	agaagtata	gctgtcgtg	gcggcagcag	tatatctga	aatgtctcat	1860
agatatata	ttttgaataa	agatggtgt	gttgaacaaa	aaaaaaaaa	aaaaaaaaac	1920
tcgagggggg	gcccgttacc	caatttcg				1947

<210> 27

<211> 3379

<212> DNA

<213> Homo sapiens

<400> 27

atgatattca	agtcgaattc	ccctcactaa	agggaaaaaa	gctggagctc	caccgcgggtg	50
gcggccgctc	taaacctact	ggatcccccg	ggatgcagga	attcggcacg	aggtctcggc	120
tccctccggc	acccgcccgg	cccggtctgt	cccggtccct	cccggccatg	gggagctgcg	180
cgcggtctgt	gctgctctgg	ggctgcacgg	tggtggccgc	aggactgagt	ggagtagctg	240
gagtgaattc	ccgctgtgaa	aaagcctgca	accctcggat	gggaaatttg	gctttggggc	300
gaaaactctg	ggcagacacc	acctgcggtc	agaatgctac	cgaactgtac	tgcttctaca	360
gtgagaacac	ggatctgact	tgctggcagc	ccaaatgtga	caagtgcaat	gctgcctatc	420
ctcacctggc	tcacctgcca	tctgccatgg	cagactcacc	cttcgggttt	cctgcacat	480
ggtggcagtc	tgcgaggat	gtgcacagag	aaaagatcca	gttagacctg	gaagctgaat	540
tctacttcac	tcacctaat	gtgatgttca	agtccccag	gcgggctgcc	atggtgtctg	600
accgctccca	ggactttggg	aaaacatgga	agccttataa	gtactttgct	actaactgct	660
ccgctacatt	tggtctgga	gatgatgttg	tcaagaaagg	cgctatttgt	acttctaaat	720
actccagtcc	ttttccatgc	actggaagga	aggttatttt	caaagctttg	tcaccaccat	780
acgatacaga	gaacccttac	agtgccaaag	ttcaggagca	gctgaagatc	accaacctcc	840
cgctgcagc	tgctgaaacg	acagtcttgt	ccctgtcaga	gaaatgacct	gaacgaagag	900
cctcaacatt	ttacacacta	tgccatctat	gatttcattg	tcaagggcag	ctgcttctgc	960
aatggccacg	ctgatcaatg	catacctgtt	catggcttca	gacctgtcaa	ggccccagga	1020
acattccaca	tggtccattg	gaagtgtatg	tgtagcaca	acacagcagg	cagccactgc	1080
cagcatctgt	ccccgttata	caatgaccgg	ccatgggagg	cagctgatgg	caaaacgggg	1140
gctcccaacg	agtgcagaac	ctgcaagtgt	aatgggcatt	ctgataacct	tcacttcgac	1200
gttaatgtgt	gggaggcatc	agggaaatcg	agtgggtgtg	tctgtgatga	ctgtcagcac	1260
aacacagaag	gacagtattg	ccagagggtg	aagccaggct	tctatcgtga	cctgcggaga	1320
cccttctcag	ctccagatgc	ttgcaaaccg	tgcttctgcc	atccagtagg	atcagctgtc	1380
cttctctgca	actcagtgac	cttctgcgac	cccagcaatg	gtgactgccc	ttgcaagcct	1440
ggggtggcag	ggcgacgttg	tgacagggtg	atggtgggat	actggggctt	cggagactat	1500
ggctgtcgac	catgtgactg	tgacgggagc	tgtagcccta	tcaccggaga	ctgcatcagc	1560
agccacacag	acatagactg	gtatcatgaa	gttcttgact	tccgtcccgt	gcacaataag	1620
agcgaaccag	cctgggagtg	ggaggatgcg	cagggttttt	ctgcacttct	acactcaggt	1680
aaatgcgaat	gtaaggaaca	gacattagga	aatgccaaag	cattctgttg	aatgaaatat	1740
tcatatgtgc	taaaaataaa	gattttatca	gctcatgata	aaggtactca	tggtgaggtc	1800
aatgtgaaga	ttaaaaaggt	cttaaaatct	accaaactga	agattttccg	agggaaagcg	1860
aacattatat	ccagaatcat	ggacggacag	aggatgcact	tgccaatcc	tcaatcctgg	1920
tttggaatac	cttgtagcag	gacatgagga	tataagaaca	ggcaaactaa	ttgtgaatat	1980
gaaaagcttt	gtccagcact	ggaaaccttc	tcttggaaga	aaagtcatgg	atatttttaa	2040
aagagagtgc	aagtagcatt	aagatggata	gcacataatg	gcacttggct	atgtccaaaa	2100
cacaaacttt	agagcaagaa	gacctcagac	aggaaacttg	aattttttta	agtgccaaaa	2160
catatagaaa	tggtttgaatg	catgggtctt	tatctaattt	atctcttctg	gacctatgtt	2220
taaatacagt	ttattttcat	gaagagaaat	gaaaaccctt	acactgatat	ctgttttcta	2280
tgggactgat	tctgaaattc	ttaactatta	agaatatatt	aatagcagca	tgacatttag	2340
cagtaatcca	ttaagggcag	tacctctaac	aaggacgcct	tccagcttca	gctatgttac	2400
ttacgtttga	tgctacttaa	agtaatgaat	gacgttttaa	ggaatcccta	accctactat	2460
cagaaaaggt	gtttgttaaa	gagccttctc	ttgtgtgtta	cgcatgaact	ttggtctgta	2520
ggtgttaaat	ggacctcttc	catgtgtata	tagtatttcc	ttgtataaag	cactttacta	2580
cctaaccttt	gtgtttgtgaa	cgtttggtga	ctgctgttga	aagaaggaaa	aggggtgtgtg	2640
agaaagccta	ctgaagcagc	agcactgccg	ctacatgttg	acaaaagtga	acatatataaa	2700
gaagtgtgtc	tatttaactc	tgaatacttg	gagaaactag	gtgaagatgc	aaccagaaaag	2760
gagaatatgt	atgcgtgaag	tctcagcttt	gagctggagg	ctagattcca	agatgacagc	2820

catgatgaaa	cttttttaaaa	aactaaacca	gaagagactt	taaaataaga	gaaagaaatc	2880
ataaatgtag	acatatgctt	ggctaaaagg	gaaatggact	ttaaatttta	aagagctcat	2940
ttgcaatgca	cttgtatata	cttcaaaaat	tattgtagac	acagaatttg	ttatatTTTT	3000
gtgcttagta	tttaaacctg	aacattgaaa	cagtttttct	ccttgtcttt	cttaacagta	3060
atagtcatta	tatttacctg	ttttttaaca	caatgtatgt	gatagtcaaa	aaatcacagt	3120
ttttcattat	tattcatctt	ctgtaccac	gcataaccac	tatacatagt	ttcttttgta	3180
cttgaatata	caaaacatga	acacagtgcc	aratgaataa	ttcacatac	agaacctttt	3240
tttctctgaa	gtcctgtgga	cttgcaataa	tatatatata	ttgctttgtt	aatttgtttt	3300
tatatttcat	atatgtaata	aaggaatatg	atctgaaaaa	aaaaaaaaaa	aaaaactcga	3360
ggggggggccc	gtacccaaa					3379

<210> 28

<211> 2006

<212> DNA

<213> Homo sapiens

<400> 28

cccacgcgtc	cgcgagcgcg	tgggattttc	tcttaagaaa	aaaaaaggaa	gacgttttaa	60
taggaaaaga	aaaacaagaa	gcagtagcag	ctaggtaatt	taaacgtctt	cctttttttt	120
tcttaagaga	aaatggaaca	tttaggttaa	atgtctttta	attttaccac	ttaacaacac	180
tacatgcccc	taaaaatatat	ccagtcagta	ctgtatttta	aaatcccttg	aaatgatgat	240
atcaggggta	aaattacttg	tattgtttct	gaagtgtgct	cttgaanaact	actgtttgag	300
cactgaaacg	ttacaaaatgc	ctaataggca	tttgagactg	agcaaggcta	cctgttatct	360
catgaaatgc	ctgttgccga	gttattttga	atagaaatat	tttaaagtat	caaaagcaga	420
tcttagttta	agggagtgtg	gaaaaggaa	tatatattct	tttttcttga	ttctgtactc	480
aacaagtctt	gatggaatta	aaatactctg	ctttattctg	gtgagcctgc	tagctaatat	540
aagtattgga	caggtataata	tttgtcatct	ttaataattag	taaaatgaat	taagatatta	600
taggattaaa	cataatttta	tacggtttag	actttatttg	ccgacctaaa	tttatagcgt	660
gtggaaattg	agaaaaatga	agaaacagga	cagatatatg	atgaattaaa	aatatatata	720
ggtcactttt	ggtctgaaat	ccctgagggt	tttttaacct	gctacactaa	tttgtacact	780
aatttatttc	tttagtctag	aaatagtaaa	ttgtttgcaa	gtcactaata	atcattagat	840
aaattatttc	cttgcccata	gccgataatt	ttgtaatcag	tactaagtgt	atcgtatttt	900
ttgccacttt	ttcttcagat	gattaaagta	agtcaacagc	ttatttttag	aaactggaaa	960
agtaatatgg	aaagagattt	cactattttg	ttcatcagtg	gtaggggggc	ggtgactgca	1020
actgtgttag	cagaaattca	cagagaatgg	ggattttaagg	ttagcagaga	aacttggaaa	1080
gttctgtgtt	aggatcttgc	tggcagaatt	aactttttgc	aaaagtttta	tacacagata	1140
tttgtattaa	atttggagcc	atagtcagaa	gactcagatc	ataattggct	tattttttct	1200
tttccgtaac	tattgtaatt	tccacttttg	taataatttt	gatttaaaat	ataaatttat	1260
ttattttatt	ttttaatagt	caaaaatctt	tgctgttgta	gtctgcaacc	tctaaaatga	1320
ttgtgttgct	tttaggattg	atcagaagaa	acactccaaa	aattgagatg	aaatgttggt	1380
gcagccagtt	ataagtaata	tagttaacaa	gcaaaaaaag	tgctgccacc	ttttatgatg	1440
atcttctaaa	tggagaaaca	tttggctgca	tccacataga	cttttatgtt	ttgttttcag	1500
ttgaaaactt	gcctcctttg	gcaacattcg	taaatgaagc	agaatttttt	ttctcttttt	1560
ttccaaatat	gttagttttg	ttcttgaag	atgtatcatg	ggtattgggt	ctgtgtaatg	1620
aacaacgaat	tttaatttag	atgtggttca	gaatatacaa	tgtaggtttt	ttaaaaagta	1680
tcttgatggt	tcttttctat	ttataatttc	agactttcat	aaagtgtacc	aagaatttca	1740
taaatttggt	ttcagtgaac	tgctttttgc	tatggtaggt	cattaaacac	agcacttact	1800
cttaaaaaatg	aaaaatttctg	atcatctagg	atattgacac	atttcaattt	gcagtgtctt	1860
tttacttgga	tatattaacg	ttctctgaa	tggcattgat	agatggttca	gaagagaaac	1920
tcaatgaaat	aaagagaata	tttattcaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1980
aaagggcggc	cgctctagag	gatccc				2006

<210> 29

<211> 3070

<212> DNA

<213> Homo sapiens

<400> 29

gttggagcct	tagaccttcg	aagtgcaccta	tagaagggtac	gcctgcaggt	accgggtccgg	60
aattccccggg	tcgacccacg	cgtccgcacac	gtatgaggag	tctgatttct	tcacatcctt	120
gtcaacatctt	gttgttgttg	ttgttgttgt	tgtttttaat	tttagctatc	ctagtagatg	180
tgaagtggta	tctcgtgctt	tttatttgca	tttctctaata	gactagtgat	gttgaacatc	240
ttttcatgtg	cttgttggcc	attcgtatat	cttcttggag	aaatgtctat	taagtccttt	300
gccaatttct	aaattggcct	gtctttttgt	tgagttataa	ccattgaatt	tttaattgctg	360
taaaacactc	catgagtaca	tgtagcatgc	actcaataac	tattagctaa	tactatagtt	420
gtaattatctt	ttgccacatg	tggcctgaat	tggaaaggaaa	ttgcctagtt	tttccaagta	480
ggagtggaga	tgtaacaaaa	cctgtattta	aggccaataa	agatttgact	ctgtcagtg	540
tgatactttg	cttttaattg	atgctcccaa	atcaagaaat	tagaaacgtg	tatacataag	600
taactagttta	tctttggcag	aaatcagtat	attatcaatg	tttaattaat	tcctaagaat	660
aataggcaat	agtgcctcta	gttttagctg	agatttagat	catctttctt	cattctgccc	720
tgttatgcga	catttgctat	aaatgccaca	atgagaattt	tttgaaacct	gggtcagaga	780
agcctacttt	ttatatctta	taagtatgtg	attgatgtat	tatcattggg	ttatgcttat	840
tgcactcttg	ttttatgggg	tgttcaaaaag	aagatagatg	agaatagggtg	atccaaagta	900
atrtagaatt	ttatcctcca	gttgggactg	aaccatcttg	aactgatgaa	tatattctac	960
cagggtttata	ggtagatatg	ccgtttcatt	atatcgggtga	tttggagtga	gacagttagct	1020
tttaagcaga	acaacagcca	tatgtaagt	taaattctgg	tattttcttt	gaataagcaa	1080
tggaaactgt	atgcattcag	gaactgggtga	ttcagatata	ttctgaaggc	cctctctccc	1140
ctgtgctccc	acagcttttg	tcctgaactc	tcctgtctcc	ttgcgatgcc	tcctgtgcta	1200
tacctatctt	gtgaactctg	cctgcaatgc	cgtcccaagt	cagttgggtga	ctcctcaaac	1260
ctcactgaag	tctttcgggg	tggattccaaa	tgtctgtttt	agccattaaa	atctgttaca	1320
tctgttacat	actttcacaa	tagcacttat	cctagtgtag	ttttaaaaat	gcacttaaaa	1380
atrttttaag	cagtgcatgc	atgcagtaaa	caaaatttgg	cagtacagaa	ataaatgggtg	1440
aagacagtct	cccagttttc	ttcccatggg	atagcacttt	ctctttttta	tccttccaaa	1500
gatattttgt	gctgatacaa	gaatttttat	accgtatgaa	gtgcttattc	atcttatctt	1560
ttacttcagg	gtaatttatg	agtaatgaaa	aaattcaagc	attacaaaaa	tgtagagtgt	1620
gtgaagcaaa	agggtccctt	tactttgatc	cctaacacta	gccctttcca	aataacactg	1680
tcctaattaa	cattcctctc	tcacagaagta	acctctgtta	taagccttgt	ctgtgccttc	1740
agattgttct	gcttacactt	atatgcctat	ttaaaagtgt	atatttttgt	atrttggttg	1800
tttgtttgtt	tgagacagag	tttctgtctt	gttgccaggg	ctgaagtgca	atggcactat	1860
ctctgtctac	tacaacctgt	gtctcccttg	ttcaagtgtat	tctcctgggt	cagcctcctg	1920
agtagctagg	attacaggtg	cacgccacca	tgcccggtta	atrtttgtat	ttttagtaga	1980
gatggagtct	caccatgttg	gccaggctgg	tcacgaactc	ctgacctcag	gttatccacc	2040
caccttggtc	tttcaaaactg	ctgggattac	aaggcatgag	ccactgtgct	cggccaagaa	2100
tttttttatc	gataccatag	tgagctctct	gcctctctgg	aacgatgtcc	actttgctta	2160
tgtatcaacc	aagcaggact	cttctctctc	tggacgcctc	ttccctgggt	tggaaatctc	2220
cagttctgtc	agaatttggt	tttcccatg	gctgcaaaact	tcacagttaa	cccttttttc	2280
tgtgtggccc	ctggggctgc	gagacaaaaa	tcctatgagt	ctgtgtaccc	tagacctttg	2340
gaagggtgaga	gcagggtccct	gagaaaaagg	aaaacgcctc	ctctccctga	tgataccctt	2400
acccgcccct	ctcctcacca	gaattgtcag	tggcctttca	ccacagtggg	ccttctctgc	2460
tgagccctgc	actgtccctg	accacacaga	agtcctggta	cctctggggc	cctgggatgg	2520
tcaccgaaga	gaagcacgct	gtccctgtct	tctcctgggt	tctgccagaa	aatcgaacaa	2580
gtgcaattaa	cacactgtta	ctgccgaagg	ctgaaactcc	caggacttgt	ccttgatcct	2640
tcacagaaacc	accaggtccg	gcacttggag	ccccccggag	agggacctcc	cagccgagcc	2700
ctcaaagaac	tcctatgaaat	caggaaactgc	ttgatgaaat	gtatctcctt	gtacctggaa	2760
gatgaagccc	aaacacccac	acctctgtct	ccccccaggg	tcgggatgtc	tcacagagcc	2820
cggccacgca	gcttcccagg	tgggtctcgg	gaggtgggag	cagggacctt	ctctgtcccc	2880
tcacacctca	ctccatccac	ctcggagacc	acctcccccc	agccagatac	ggaataaaac	2940
tacagacgca	gaaaaaaaaa	ggcgggccgc	tctagaggat	ccctcgaggg		3000
gccccagctt	acgcgtgcat	gcgacgtcat	agctctctcc	ctatagttag	tcgtattata	3060
aggagcaaaa						3070

<210> 30

<211> 2227

<212> DNA

<213> Homo sapiens

<220>
 <221> SITE
 <222> (289)
 <223> n equals a,t,g, or c

<400> 30
 cggacgcgtg ggtcgaccca cgcgtccggg aaaaarggaa aaratgccgt gtaaaatctc 60
 gttctgtgtc tgaattgccg taggctcaga tcttcatttg aggttctgtg tctgaattgc 120
 cgtaggctca gatcttcatt tgaggttatg ttctataagt taacgttgat cttgtgtgag 180
 ctttcggtag ctggagtaac acaggcggcc tcacagcgac ctctccagcg ccttccaagg 240
 cacatctgca gccagcgtaa tctcctggg agatgcctcc tcaaggccnt gctccagacc 300
 acgtggggar ggcctgacar ccaattccca ggctgtcccc acccttgrag agtgacccta 360
 aacgctagac agatggggaa tgggaaagaa aagaaagctg cagacctcaa gttaaaattc 420
 cctcaaaaac gtttttattt atctgctttt tctgaaagga taaaggctct ttgaaaatta 480
 ttttctaaca aataacatga acacttctag aaaccctaga aaaacacaaa gtattcaaaa 540
 tagaaagaaa aattacccat tactctttaa gccagcatta tccattgcgg tgcttttgga 600
 gttgggtgag gccgtagcct ctgccaaagc aaggagcccg gtggtggctg tggcattcct 660
 gcagggttgt tttttttct ttgagatgga gtctcactct tgtcacccca gctggaatgt 720
 ggtgggtgtaa acagctcact gcagccttga ccctgaggct caagcgatcc ttctgccttg 780
 gcctcctgag tagctgggat cccaggcgag agtcaccaca ccctgtccat gttcctgcag 840
 gtcttgatat gcgaggacgc tgtgtcttcc ctgccacatt ttcttcttct ttcttgagac 900
 agacccttgc tccatcacc cggccagagt gtggtsgtgc gaacacggct cactgcagcc 960
 tcgaccctca ggctcaagcg atcctcacgc ctgggacccc caaagtgtcg ggatcacagg 1020
 cgagagtcac catgctggcc tgaatcttca gggatatttta cgggtgaagt gtcacttact 1080
 tarccatscc tgtttcaaga gtgtaggttg tcaccctgtc tctgycgctg acctggcctg 1140
 gaccctcgcc tgtgagaggg aggggtgggc tgggctggag gaacctraag cctcgtgat 1200
 gtcacaagcc catctggctg ggcattccct gctgtgtcct gagctgcaca tgccccagg 1260
 ggccccaca gcagaggcga gccactgrag ggtgragggc ttccacggac ggtcttcagg 1320
 ggragaagaa gggcccaggc ccccaggaga ctgaggagac cagagcctgg ggtcaggggc 1380
 tyagcagggg ct yarccagg gctggatgtc cggagccagc cccgmagccc tgkgtcttt 1440
 gttcttcgca ctcccaccgt ccgtgtgaac agctccagcc ccacctgcgc ctccctgtgc 1500
 tgggctccat cagggagccc agaagacgtg tgtgcttctg aaattgggtc cctacatgcc 1560
 tttgtcccag tgcaccttgc tcttccatt tactatcgag atttaaattgc ctgttttctc 1620
 cccagagggt gacggatata ttacagcgtt acgacacgga tcaggacggc tggattcagg 1680
 tgtcgtacga acagtacctg tccatggctc tcagtatcgt atgaccctgg cctctcgtga 1740
 agagcagcac aacatggaaa gagccaaaat gtcacagttc ctatctgtga gggaaaggag 1800
 cacaggtgca gttagatgct gttcttctct tagattttgt cacgtgggga ccagctgta 1860
 catatgtgga taagctgatt aatgggtttg caactgtaat agtagctgta tcttctaat 1920
 gcagacattg gatttgggtg ctgtctcatt gtgccatgag gtaaatgtaa tgttccaggc 1980
 attctgcttg caaaaaaatc tatcatgtgc tttcttagat gtctctggyt ctatagtga 2040
 aatgctttta ttagccaata ggaattttaa aataacatgg aacttacaca aaaggctttt 2100
 catgtgcctt acctttttaa aaaggagttt attgtattca ttggaatatg tgacgtaagc 2160
 aataaaggga atgttagacg tgtaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2220
 aaaaaaa 2227

<210> 31
 <211> 1288
 <212> DNA
 <213> Homo sapiens

<400> 31
 cccccgggct gcaggaattc ggcacgagcc tgacctcccc agcctcatct ctccctcctc 60
 tgtctctgcc ctctgtgtc cagccaacgt ggctgtcac tcgtccacct gccatactgt 120
 cctgactcca ggcttttgcc tgtgtatgg cctctgttgg gaccactctt gtctctcccc 180
 tgtgtgtct gctaattccc atcgtgtca gtgatccat gctgcagaac acaccactcc 240
 atccatggaa aacaatcaca atcattgatt actatctctc cctgggcttc ctgggggtgga 300
 ctgggctcag ctgggtgggt cactttgggg cctcagcagt catgggcaga cagtggctgg 360

ggtcactgca	aagactcccc	tgcatctctg	gcagttgatg	ctggctgtca	tctgagacac	420
ctacccaggg	cctctccctg	gggcctgggc	tcctgcttag	cttggttggg	aggctccaag	480
accaacatcc	caagaaagat	gagacagaag	ccagatcacc	tttttgggcc	tggttcaga	540
agtcaccgag	catcacttct	gctgcattta	tttcttaaaa	cacaaatata	aaaccccatc	600
tcttgatggg	agggggcctc	atggtttgta	aacatgttct	aaactccact	ctgcccggcc	660
ttggctcaac	gtgtctggta	atgtgtgggc	tgtgaggctc	cccgaacgta	gacctcagac	720
tgcaacgctg	gccgttacag	ggtctggcac	acgggcccac	gtcaggccca	tcgccacagt	780
gatggttggt	ctgtgactgt	ttctgggtggc	ctctgctcca	cactccaggc	tgacgctgtg	840
ccccctccac	tgggaccctc	gggtggcttc	catgcacttg	tgccctaaat	cctgtctcta	900
gactaaactt	catctcctgt	gttctcattc	tgacagatgg	ctgttaggga	acctgaccat	960
ctgcagcgcg	tctcgttgcc	aaggataaat	gtcagtgcc	cccttcagtg	gctcccatgt	1020
cacagaattg	tcctgcagcc	ctggcacatg	tgtgccatgt	gggagctggg	gcaggtcctc	1080
tttcatcctg	tggctccgag	ggagggggcc	gctccttccc	cagctcttac	cctgacttgg	1140
ccctcgctct	gcagccactc	agagagcacg	atggagctgg	agcttcagtt	ttgaccaa	1200
gcgtgtcgcc	ggcttttgtg	tgtgtgtgtg	tgtgacagag	ccagaccctg	tctttaaaaa	1260
aaaaaaaaaa	aaactcgagg	ggggggccc				1288

<210> 32

<211> 3280

<212> DNA

<213> Homo sapiens

<400> 32

gcgtccgggtg	agaggcaagg	acttttcatc	tttagggcat	tctgaccacg	tcctgtttca	60
gagagattgt	ccccggcgtc	tcagtgttat	ggggagcagg	ttcctcctgg	tcctgtcttc	120
agggtctcact	gtcttactgg	ctctgccagg	atcagaagcc	aagaattctg	gagcttcctg	180
tcctccatgc	cctaaatatg	ccagctgcc	caacagcacc	cactgtactt	gtgaagatgg	240
ctttcggggc	agggtctggca	ggacatactt	tcattgattcc	tctgagaagt	gtgaagatat	300
taatgaatgt	gaaaccgggc	tggcaaatgt	caagtataaa	gcatattgta	ggaataaagt	360
tggagggttac	atctgtagct	gtttggtaaa	atatacttta	ttcaactttc	tggctgggtat	420
tatagattat	gatcatccgg	attgttatga	gaacaatagt	caagggacga	cacagtcaaa	480
cgtggatatt	tgggtgagtg	gggtgaagcc	tggatttggg	aaacagctgg	tacgtataac	540
tatgccattt	tcctacccaa	acattaacat	gtcttcctgt	gatttttagg	gtagggtagt	600
tctatccagg	ggtaattttg	tcctctgtcc	caaggtcatc	tgtcaatgac	tggggacact	660
tttggttgtc	ataccttggg	ggtgatgtgt	gtgactggca	tctggtggat	ggagaccagg	720
gatacagctc	aacatcctac	agtgccagg	acagcctccc	acaatcaaga	agtgcctagt	780
gccatagtgc	catagagata	gagaaataca	agtgtagggg	gaaagtgcct	cagctggcat	840
tcaaagacct	acatcaagca	cctagattct	caatgccaca	cgcaccttgt	agcaccttaa	900
taaatatcgt	tgccttctgg	gtccccact	ccaacacttg	tgcataatcc	ctatttctta	960
catttcagta	agagtcaatc	taattagctt	aatttttttg	gaaggaaaat	ctgagaagaa	1020
atggaagcag	agaggacttt	gcaagaaggg	ctactcaact	aattcaaagc	gtggagttga	1080
gcatctggaa	tgcgagtttt	gcttctccag	gaaagggtca	aatttctgaa	tttgatatag	1140
tctatgaaac	caagaggtgc	aatgagacaa	gggagaatgc	ttttctggaa	gctggaaata	1200
acaccatgga	tatcaactgt	gctgatgctt	taaaaggaaa	cctaagagag	agcactgcag	1260
ctgccctatc	acttatcaat	ctcttgggga	tattctgaat	gcatcctttt	ttagtaaacg	1320
aaaagggatg	caggaaagta	aactgaactc	ttacgttgtg	agcggcacca	tcggtttgaa	1380
ggaaaaaatt	tcctctctctg	aacctgtgtt	cctgactttt	cgccataatc	agcctgggtga	1440
caagagaaca	aaacatatct	gtgtctactg	ggaggagatca	gagggaggcc	gctgggtccac	1500
ggagggctgc	tctcatgtgc	acagcaacgg	ttcttacacc	aaatgcaagt	gcttccatct	1560
gtccagcttt	gccgtcctcg	tggctcttgc	ccccaaaggag	gacctgtgtc	tgaccgtgat	1620
caccaggtg	gggctgacca	tctctctgct	gtgctcttct	ctggccatcc	tcaccttctt	1680
cctgtgccgg	ccatccagag	acaccagcac	ctccctccat	ctagagctct	ccctctgcct	1740
cttcttgccc	cacctcctgt	tcctgacggg	catcaacaga	actgagcctg	agggtgctgtg	1800
ctccatcatt	gcagggtctg	tgcaacttct	ctacctggct	tgcttcacct	ggatgtctct	1860
ggaagggtcg	caggtcttcc	tcaccgtcag	gaacctcaag	gtggccaact	acaccagcac	1920
gggcagattc	aagaagaggt	tcattgtaccc	tgtaggctac	gggatccag	ctgtgattat	1980
tgctgtgtca	gcaatagttg	gacccacagaa	ttatggaaca	tttactcact	gttgggtcaa	2040
gcttgataaa	ggattcatct	ggagcttcat	ggggccagta	gcagtcatta	tcttgataaa	2100

cctggtgttc	tacttccaag	ttctgtggat	tttgagaagc	aaactttcct	ccctcaataa	2160
agaagtttcc	accattcagg	acaccagagt	catgacattt	aaagccattt	ctcagctatt	2220
tatcctgggc	tgttcttggg	gccttgggtt	ttttatggtt	gaagaagtag	ggaagacgat	2280
tggatcaatc	attgcatact	cattcaccat	catcaacacc	cttcagggag	tgttgccttt	2340
tgtggtacac	tgtctcctta	atcgccaggt	tcgaatggaa	tataaaaagt	ggtttagtgg	2400
gatgcggaag	ggggtagaaa	ctgaaagcac	tgagatgtct	cgctctacta	cccaaaccac	2460
aacggaagaa	gtggggaagt	cctcagaaat	ctttcataaa	ggaggcactg	catcatcatc	2520
tgcagagtca	accaagcaac	cgcagccaca	ggttcatctc	gtctctgctg	cttggctaaa	2580
gatgaactga	cctggcaagt	gccatggcaa	tgacccggaa	gttacccttc	ctttccgttt	2640
gtctacagcg	cccctgtggt	cacacataga	ttggacaaat	gccactattt	ctagctttcc	2700
tgtgaaaagt	ctaggctcat	tcacctattt	tggcttttta	tgttcataga	aagaacaaga	2760
catttgggag	aattcttaga	tccagagtcc	agttagtggtg	cacgtgcaat	gaagtgtcgg	2820
aaggatgcat	tttaaagatg	gcgggcgggg	gaagtggatt	ttcttcttgc	agctactgcc	2880
accttgccag	aaactcacta	actggcatct	ggattcagct	catagttccc	tttctggcct	2940
ctctgctgta	ttttatgctc	ccaaagatct	tacattaaca	ctccacattc	acataattca	3000
acaattttca	tatggatcag	tattaaagag	ggtgttgcat	tttgcaatac	aaaaatgcat	3060
tatcaggtgc	tggagaggat	gtggagaaaat	aggaacactt	ttacactgtt	ggtgggactg	3120
taaactagtt	caaccattgt	ggaagtcagt	gtggcgattc	ctcagggatc	taggactagg	3180
aataccattt	gacacagcta	tcccattact	gggtatatac	cccaagggct	ataaatcatg	3240
ctgctataaa	gacacatgca	cacgtaaaaa	aaaaaaaaaa			3280

<210> 33

<211> 1297

<212> DNA

<213> Homo sapiens

<400> 33

ccacgcgtcc	ggactgcttt	acggacattg	gatgaagccg	aagcatttag	aatggtgcct	60
ggcacacagt	tgggtgcgtga	tatggttaag	ctttgtgtcc	ccaccacat	ctcatcttga	120
atgtgacggt	ttcccgggt	ccctcctgcc	gccatgtgaa	gaaggctgtt	gcttccccct	180
caccttccac	caccatgatt	gccatggatg	ctccccactc	caaagcagcc	ctggacagca	240
ttaacgagct	gcccagagaac	atcctgctgg	agctgttcac	gcacgtgccc	gcccgcagc	300
tgctgctgaa	ctgccgcctg	gtctgcagcc	tctggcgagg	ctcatcgac	ctcatgaccc	360
tctggaaaac	caagtgcctg	cgagagggct	tcatacccaa	ggactgggac	cagcccgtgg	420
ccgactggaa	aatcttctat	ttcctacgga	gcctgcatag	gaacctcctg	cgcaaccctg	480
gtgctgaaga	ggatatgttt	gcattggcaa	tgatttcaa	tgggtgggac	cgttggaagg	540
tggagagcct	ccctggagcc	cacgggacag	attttcctga	ccccaaagtc	aagaagtatt	600
ttgtcacatc	ctacgaaatg	tgcctcaagt	cccagctggg	ggaccttgta	gcccaggggct	660
actgggagga	gctactagac	acattccggc	cggacatcgt	ggctaaggac	tggtttgcctg	720
ccagagccga	ctgtggctgc	acctaccaac	tcaaagtgca	gctggcctcg	gctgactact	780
tcgtgttggc	ctccttcgag	ccccacctg	tgaccatcca	acagtggaac	aatgccacat	840
ggacagaggt	ctcctacacc	ttctcagact	accccggggg	tgtccgctac	atcctcttcc	900
agcatggggg	cagggacacc	cagtactggg	caggctggta	tgggcccaga	gtcaccaaca	960
gcagcattgt	cgtcagcccc	aagatgacca	ggaaccaggc	ctcctccgag	gctcagcctg	1020
ggcagaagca	tggacaggag	gaggctgcc	aatcgcccta	ccgagctgtt	gtccagattt	1080
tctgacagct	gtccatcctg	tgtctgggtc	agccagaggt	tccctccaggc	aggagctgag	1140
catgggggtg	gcagtgaggt	ccctgtacca	gcgactcctg	cccgggttca	acctaccag	1200
cttgtggtaa	cttactgtca	catagctctg	acgttttgtt	gtaataaatg	ttttcaggcc	1260
gggcacagggc	ggccgaggat	gtgctgctgg	ccgctgctcc	tgtgtggggg	gctgctcccc	1297

<210> 34

<211> 2184

<212> DNA

<213> Homo sapiens

<400> 34

ggcacgaggc	ggccgaggat	gtgctgctgg	ccgctgctcc	tgtgtggggg	gctgctcccc	60
------------	------------	------------	------------	------------	------------	----

gggacggcgg	cggggggctc	gggcccgaacc	tatccgcacc	ggacccctcct	ggactcggag	120
ggcaagtact	ggctgggctg	gagccagcgg	ggcagccaga	tcgccttccg	cctccaggtg	180
cgcactgcag	gctacgtggg	cttcggcttc	tcgcccaccg	gggccatggc	gtccgcccac	240
atcgctcgtg	gcgggggtgg	ccacggggcg	ccctacctcc	aggattatct	tacaaatgca	300
aataagagagt	tgaaaaaaga	tgctcagcaa	gattaccatc	tagaatatgc	catggaaaat	360
agcacacaca	caataattga	atctaccaga	gagctgcata	catgtgacat	aaatgacaag	420
agtataacgg	atagcactgt	gagagtgtac	tgggctctacc	accatgaaga	tgcaggagaa	480
gctggctcca	agtaccatga	ctccaatagg	ggcaccaga	gtttgcggtt	attgaatcct	540
gagaaaaacta	gtgtgctatc	tacagcctta	ccatactttg	atctggtaaa	tcaggacgtc	600
cccatcccaa	acaaagatac	aacatattgg	tgccaaatgt	ttaagattcc	tgtgttccaa	660
gaaaagcatc	atgtaataaa	ggttgagcca	gtgatacaga	gaggccatga	gagtcctggg	720
caccacatcc	tgctctatca	atgcaacca	aactttaacg	acagcgttct	ggagtcgggc	780
cacgagtgtc	atcaccctaa	catgcccgat	gcattcctca	cctgtgaaac	tgtgattttt	840
gcctgggcta	ttggtggaga	gggcttttct	tatccacctc	atgttggatt	atcccttggc	900
actccattag	atccgcatta	tgtgtcctta	gaagtccatt	atgataatcc	cacttatgag	960
gaaggcttaa	tagataattc	tggactgagg	ttattttaca	caatggatat	aaggaaatat	1020
gatgctgggg	tgattgaggc	tggcctcttg	gtgagcctct	tccataccat	cctccagggg	1080
atgcctgagt	tccagtctga	gggtcactgc	accttggagt	gcctggaaga	gctctggaag	1140
ccgaaaagcc	aagtgggaatt	catgtgtttg	ctgttcttct	ccatgtctac	ctggctggca	1200
gagcatcagg	ctgcgtcatt	ttcgaaaagg	gaaggaaatg	aaattacttg	cctatgatga	1260
tgattttgac	ttcaatttcc	aggagtttca	gtatctaaag	gaagaacaaa	caatcttacc	1320
aggagataac	ctaattactg	agtgtcgtca	caacacgaaa	gatagagctg	agatgacttg	1380
gggaggacta	agcaccagga	gtgaaatgtg	tctctcatac	cttctttatt	acccaagaat	1440
taatcttact	cgatgtgcaa	gtattccaga	cattatggaa	caacttcagt	tcattggggg	1500
taaggagagt	tcagagccag	tcacgacctg	gcctttcatt	atcaaaaagc	ccaagcaata	1560
taaaaacctt	tctttcatgg	atgctatgaa	taagtttaaa	tggactaaaa	aggaagggtc	1620
ctccttcaac	aagctgggtc	tcagcctgcc	agtgaatgtg	agatgttcca	agacagacaa	1680
tgctgagtgg	tcgattccaa	ggaatgacag	cattacctcc	agatatagaa	agaccctata	1740
aagccagaac	cctttgggtg	gtggcacgtc	ttcttccctc	ttcccttgcc	acagagattt	1800
ctcccatcca	acttgcttgt	ttgccttctg	ctaactcagc	tgacagctga	gcaccaagag	1860
cttgtgatca	aaattctgtt	ggacttgaca	atgttttcta	tgatctgaac	ctgtcatttg	1920
aagtacaggt	taagactgtg	gtccactttg	ggcgggagac	gtgtggagac	ttttcttccc	1980
cattttccct	ccctcctttt	tcctttccat	gttacatgag	agacatcaat	cagggtctct	2040
tctctttctt	agaaatatct	gatgttatac	atacatggtc	aataaaataa	aactggcctg	2100
acttaagata	accattttta	aaaattgggc	tgcatgtgtg	gaataaaaga	attctttctt	2160
tcctaaaaaa	aaaaaaaaaa	aaaa				2184

<210> 35

<211> 949

<212> DNA

<213> Homo sapiens

<400> 35

ggcacgagct	cccatgacc	ttattgctca	gagtgcagta	ttagggcaag	gtttcgccac	60
tgccctccctc	catgaatgta	ttcttccctc	ctgccctggg	gacatgggga	gtggcccggt	120
tctttcccca	tctagtccca	gaaagatggc	gtttgggttt	ctgttggttg	attttttttt	180
tttttttttt	tttttgacc	aaagtggcaa	ctaggtcagt	gttgggggat	caagctggcc	240
tcgggggtgg	gggccccac	ctgcctctcc	ctggttccca	cagtgttagc	gtccctgaaa	300
agacaatat	ctctctaaa	caataagggg	tgacggggcg	gggggaggtg	ttgctgctgc	360
tgccccccag	ctcccttcc	ctgccaggtg	tgggggagac	tcctgttggtg	actgaatgta	420
acccccccac	ccctgccgca	gccaatgcag	gggaaggggg	acactcttcc	tgtctcttct	480
ccccagctaa	agagactttg	gacttagggg	gccccatgagc	ctggagaggg	cttaaccttg	540
tgaggaaagta	taggggggagc	cctctccccc	ccccatcccc	ttctgagagt	ggtcaatgtt	600
tacaagcccc	tgagcccccc	tgcccaggga	ctcagacccc	gttgctgtcc	ttccccggcc	660
cgggtcttcc	tgggccccctg	ctgctctctc	gccctttctg	gggttggggg	gggtgcaggg	720
gtcaccgtgt	tcctgtctgt	ccttgtacc	acagtcctcc	cgtccctctc	ccacctgtg	780
tgacttccct	ctcttttacc	tgctcctgta	aatactccct	tctcccaata	aaacttggtg	840
tgtgttctcc	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	900

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

949

<210> 36
<211> 3338
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (2861)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (3328)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (3330)
<223> n equals a,t,g, or c

<400> 36
taccgggaca ttctggagtc cgagaagtc aacggcgcggt tgctgcggcc gccgcgctcc 60
ccggcccgag gcgatggaga acggagcggg gtacagcccc actacggagg aggaccggg 120
ccccgccaga gggccccgga gcggcctcgc tgccactctt ttcattgggccc ggctccatt 180
gctccggcgc gttctcaagg gcttgagct gttgctgctt ctgctggcct tcatctgtga 240
agaagtgtga tcacaatgta ctttatgtgg aggaactttat ttttttgagt ttgtaagctg 300
cagtgccttt cttctgagtc tccttataact gattgtgtat tgcactccat tttatgagag 360
agttgatacc acaaaagtaa aatcatcgga tttttatatt actttgggaa caggatgtgt 420
gtttttgttg gcatccatca tttttgtttc cacacatgac aggaacttcag ctgagattgc 480
tgcaattgtg tttggattta tagcaagttt tatgttccta cttgacttta tcactatgct 540
gtatgaaaaa cgacaggagt cccagctgag aaaacctgaa aataccacta gggctgaagc 600
cctcactgag ccacttaatg cctaaagact ctggggagca gatgttacct aaggtagtga 660
ccctgcattg tgggtgcctga gccctggcag aagctcttgt aaaatttgtt aattgtttaa 720
accacttctt ttggagagca aggggaaggt caagaaggca gttttatcaa tattgtgtca 780
gtcaccacaa agtagggcag ataagttaaa aaaaattttt ttttaataaa taattgaac 840
ttatctcaaa tggagatttt ggtgggagga ggagaaaaca attgttttta aatcacacag 900
ctcaacggtt gataaatgat tctgtcattc tgttacaggt cattctttta ctaggcttag 960
cttccaaatt atgctttata gctgtataaa catcgtgatt atattcatct acttagaaat 1020
tgttttatct ttaaatcaat ttgcttagct gtttgttttg atgcttagat tatgttctgt 1080
taatgggaat ttaacatatt taagaaacca atatttaaaa tgttgggtcta ggtttttttc 1140
cttaacatat attaccaggc ttacttgtat ttcactcagc cttaaatgtt ataatatctt 1200
tggataacgg ttattaatc tttgagacct tcgtatagcc tataaatgtt atgggagatg 1260
ttggtatttt atgtgtataa aagcaacaat atcagcaact tcgtgtttat actgcacctt 1320
ggttgttgat gtcaagtaaa aaaaagattg ttttgtaaca cataaaaaaa tggagaagaa 1380
tgataccaca cctaaggacc aaagataaga aagacttttt gcccaagaca gtgaaagtaa 1440
ttataaaaaa aagctttgac cacttaccaa gtatctgaag agatgagttc atactatgat 1500
ttagaaagtg gttcaattcc cctgttggca tatgattatt tttactaaaa ttaatacagc 1560
tctgtgggtc ttccttagtg ttttctttga agccaatctg ttttttttag gacaccagcc 1620
tttggttttt catctgttcg agatgccttc tctctgtctc cttatcagat agaaatggag 1680
tcattgtgct ctgcttcac tagcagaggt tggcctctgg ctctgacac tttgtcagt 1740
tgtctttagg tggctcctgaa tcttgggccc ttttgattgt gaatactgtg tagcaggatc 1800
ttgagagtcc ttgtttctac ataggcattg ctctagtttg tctttggcaa aaaaaaaaaa 1860
aaaaaaaaag aaatatccag ggaacctctc ccagactaat actgttgggtg gcataagaga 1920
atcaagccat tctcaagaga taacttcata accagaattg tctgttgggt agcagctgtc 1980
acagataggc agggcacttg ggatatgacc tttctgtcca ggtgattcac agactagacc 2040
tttcttatcc tccctcctaga gttttgactt gggactctag tgttaagatg atgagccctg 2100

gcatcaggtc	cttctgcact	ttggtggaag	tctcccagg	taggtttcct	atltgaaaca	2160
gtggaatcat	gtttccagtg	ataaagttaa	atgacctcat	cctttttttt	ttttctcatc	2220
tgccatttgt	gtgtcttaga	tgggttttaa	ttgcatgaat	gtggctaata	tggtttctcag	2280
aaattgggtca	gtatggcccr	acatagcttc	tgctctgtct	tactgactca	atacctttag	2340
gattttgtatc	agagtttgga	tactagtgtt	agtgggtggg	tcaccactac	tttaattggga	2400
gataatgaaa	ccaatcatgg	atgctgtttt	tattgggcat	gtcatctaag	agaggagaaa	2460
tagctgggtt	ttgggtctaa	ttatgaataa	ggactgattc	agaaaacgag	tttatggtag	2520
gtagactaaa	gtttcacatc	agactgtacc	attgtgattt	agacctatct	aaaattcaga	2580
gcataatcatc	tggtgtacct	cagggtcacc	acccatgtat	tggtgttagt	caggattgac	2640
agatacatte	tcagctggcc	tgatcatata	aacatactgt	cattgagctt	aagctccgct	2700
tggtctgagg	tttcacctcc	atgtgtttca	ttggtgcaaa	agtggatctc	ttagtgggtc	2760
acttaattct	ttctttttca	gaaagatagt	atgttcaactg	gtatatttgg	tcactcttag	2820
aaccttctct	cacattgttt	tttatgggac	ccatgaatgg	nttagccttt	cttttctatt	2880
gtagaaggaa	ataaatagga	gtaaaaagac	cattgtagta	aataaagtta	aggggaactt	2940
gggaccagaa	accactgtta	tgtacaaaaa	aatggcaaat	tcaataaact	caaatttaaa	3000
ataattttta	aattaacagt	targataaat	tttatatttt	atacaaatag	attgcttaga	3060
atgggtctca	agaattataa	gagaaatgaa	ctcacagtac	aaaaatttta	taattactat	3120
acttgtgttt	tggttggggg	ctgggaaatg	tatttttaca	ctgtagccaa	tcattttata	3180
tttgtcaatt	taaatcttat	gggtcttttt	ttttttatct	ctcttgatgt	cagattttat	3240
agtcttttta	aataaatcca	tttaattaaa	acgttaaaaa	aaaaaaaaaa	aaaaaaaaaa	3300
gggcggccgc	tcgcgaatct	agaactangn	cccacgcg			3338

<210> 37

<211> 1563

<212> DNA

<213> Homo sapiens

<400> 37

cggcacagg	agaaaaggac	tcagaccttc	cagattggct	ccagaaggca	gctgctgtca	60
gtgattttcc	ttacagacca	ttgcagccac	ctaccttgcg	ccatgggtccc	gagacagcat	120
ggcgtcttcc	cccacacgga	cctgcagggtg	gcctcttagg	agctttctcg	ctgaaggagg	180
tcataccctct	tagtcacatc	tgagcagcac	tcacttaggc	ctcggccccc	agggctcatgc	240
ggggggcacct	ggcgggttcc	ccggctctgt	ctgggtctggc	atctgtctgc	ctttgggcta	300
ccttttctggc	acagctccct	gggcctgtgg	ctgctacaag	ctggactcca	gctccccctgg	360
gctgcagtg	tgctcgcagt	ggtccagaaa	agaggctggg	cacagcggct	cctgggtcag	420
cagcatcctt	ggcgcaggct	ggcccaggag	ctccgtgcag	ggtgctgcca	gttgaccag	480
ccccggctgc	actcaatgtc	cgagaaccgg	gctgggtggg	gggcctcttt	gacggagctt	540
tactccaagt	cttgctgaat	ttcctgagga	aaagcactga	tggtctcatg	gacacaagg	600
aggcagaatc	tcttgaagta	gaataaactg	cactccttct	cagtcttctt	ttctcagctt	660
ctcctggacc	gccagctcct	ccatgcgcct	cagacactcc	caacgccgca	ctgcgggggt	720
tcctcccgcc	ctggcccatc	tcataccctt	tggtgttgga	tacagctccc	ctgtgtgcat	780
gtcgcctctgt	ggcagatgct	cagggaacttc	tcagactcca	ggatcactcc	gagtacactg	840
accacacagc	ctgtgcaca	aactgcagcg	ccagctaaag	accaggaatc	tgacattgtc	900
ggaggtgaag	ggatcctttg	tgacatagct	ttctctcagg	aggaccatcc	actgggtgtc	960
ggggggagctt	ctgccccaa	tagcaggagg	gagctgtcaa	ggcgtggagt	tcacactcaa	1020
acactcccag	aggacggcac	tctgcatggg	acgcatcca	gctcctttga	ctgtggaatc	1080
aagtacatta	tcagctggcc	cctggctcct	ggctgtgacc	tcctctcgct	tgaactgagc	1140
cttgtgtgta	agggcgtatc	aagttgtatg	ggcttcccg	ctgggtgacc	ccaaatgcaa	1200
ctccaggaag	catgggttac	caatgactaa	ggttaacgat	cagctgggcg	cgggtgctgac	1260
gcctgtaatc	ccagcaacttt	gggaggtcga	ggcgaatgga	tcacgagggtc	aggagtccaa	1320
gaccagcctg	gccaacatgg	tgaaaccccc	atctctacta	aaaatacaaa	aattagccgg	1380
gcgtggtggc	gggcgcctgt	agtcaccagt	actcaagagg	ctgaggcagg	agaatcgctt	1440
gaactcggga	ggcgcagctt	gcagttagct	gagatctcgc	ccgtgcact	ccagcctggg	1500
cggcagagcg	agactccgtc	tcaaaaaaaaa	aaaaaaaaaaa	aaaaaaaaaa	cgaggggggg	1560
ccc						1563

<210> 38

<211> 1048
 <212> DNA
 <213> Homo sapiens

<400> 38
 cccgggtcga cccacgcgtc cgacaaaaaa caaggtatgt gtgtgccttg ggatgctttt 60
 ctgggcttac cccatctgtg tcttcataga ttctctgtcc tgcagccat gtctttggte 120
 cacaggagcc acttcccact tcaactctcc cacaacctcc ccactgttta ctctcttcat 180
 gccctgtgct cttgcccaca atcccttcac ccaattggga aaactggatg atagataagt 240
 agtttcttat ttacatgcat ttctctggaa attaaagtgt ctttccagct agtatttcct 300
 cacgtcggaa gcatgtgtaa aataacacat ggaaactatt tccacctggc acataagaaa 360
 aaaaggtct gcctatatgt taactgactt ctacactttg catgttttag tctgaccac 420
 aatttttttt ttggatagt tgaagtaaaa aacgtgggta caagatgact aaagtttggc 480
 ctacaatttc ctgttttgac aacaggaaga attgtgcagt gttttgcaac aggactaatt 540
 ctgattcttc actgccctta aagataacta gagggaaagg gtcttctctt tctttttcat 600
 ttattaaaga tacttacagg ttcttcagaa gaagtgatgg gtcttttagg taatgaaata 660
 ggtgttgatg gtgttatggg tgatgatgta actgggtggc gctgtataaa gtcttcatcc 720
 ttaaacatca gaactgagag ttgggttact gattttaaag aaaatgtgta ttaatctaca 780
 gtccactacc ttgactacca ggaacttgct acaactacct catatcctgg cacaggattc 840
 aaaaggcaaa atcttgaacc taattacgtg tagccaaata ttaagcttcc tcttagatca 900
 aacagtacaa taaatatctc ttattaaagg aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 960
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1020
 aaaaaaaaaa aaaaaaaagg gcggccgc 1048

<210> 39
 <211> 1430
 <212> DNA
 <213> Homo sapiens

<400> 39
 cggcagcagg agcagctgag tcccttccct gtctttcact cttctggcat cgggtggttt 60
 acttcttcga ttgaacctg cttctctgac ccccttggga ggccgccttc ttcaggcgcc 120
 tcccttctct ccacgagctc gctctgacag ctgaggaact ggcaagatcc tgctaccag 180
 aggggtgaatg ggtatctttc ccggaataat cctaattttt ctaaggggta agtttgcaac 240
 ggccggccgtg attgtaagcg gacaccagaa aagtaccact gtaagtcatg agatgtctgg 300
 tctgaatttg aaaccttttg tatatggcgg ccttgcctct atcgtggctg agtttgggac 360
 tttccctgtg gaccttacca aaacacgact tcagggttcaa ggccaaagca ttgatgcccg 420
 tttcaaagag ataaaatata gagggatgtt ccatgcgctg tttcgcatct gtaaagagga 480
 aggtgtattg gctctctatt caggaattgc tctgcggtt ctaagacaag catcatatgg 540
 caccattaaa attgggattt accaaagctt gaagcgctta ttcgtagaac gtttagaaga 600
 tgaaactctt ttaattaata tgatctgtgg ggtagtgtca ggagtgatat cttccactat 660
 agccaatccc accgatgttc taaagattcg aatgcaggct caaggaagct tgttccaagg 720
 gagcatgatt ggaagcttta tcgatataata ccaacaagaa ggccaccagg gtctgtggag 780
 ggtaagtact cttttcctgc tattatccta cactctcagt tcttacaatt tgcagagaat 840
 ttttttttat ataaagacat aaaatcgtga attataatcc aaaaactaag gtaagaaact 900
 cctcatctcc cttgaaaggc ccaaaactta tcattggcct tttattttctg cataatgttt 960
 ggggattata taggtgggga aagtatttac attatttgag atggctgttt cgatcatatt 1020
 cacagtgaat gtagttgttc agtgtatttt ttgcaagtt ctgtactaac acgatgatgt 1080
 atgtctttgt agtgcttatg ttcaaaagct gttaccggct gtgtgctgtg gctcatgcct 1140
 gtagtcccag cactttggga ggccaacgcg ggtggatcac ttgaggtcag aagttcaaga 1200
 ccagcctggc caacctgggt aaaccccac tcaactagaa atacaaaaat tagccaggca 1260
 ttgtgtgtga tggctgtgtt cacagctact caggaggctg aggcgggaga attgtttgag 1320
 ccctggaggt ggaggttgca gtgagccaag atcatggcac tgcactccag cctgggtgac 1380
 ggggcgagac tctgtctcaa aaaaaaaaaa aaaaaaaaaa aaaactcgag 1430

<210> 40
 <211> 2103

<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (2101)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2102)
<223> n equals a,t,g, or c

<400> 40
ttcctcgtag cgagcctagt ggcgggtgtt tgcattgaaa cgtgagcgcg acccgacctt 60
aaagagtggg gagcaaaggg aggacagagc ccttttaaac gaggcgggtg gtgcctgccc 120
ctttaagggc ggggcgtccg gacgactgta tctgagcccc agactgcccc gagtttctgt 180
cgcaggctgc gaggaaggc ccttaggctg ggtctgggtg cttggcggcg ggggttctt 240
ccccgctcgt cctccccggg ccagaggga cctcggcttc agtcatgctg agcagagtat 300
ggaagcacct gactacgaat gctatccgtg cgagaacagc tattccacga gaggatccgc 360
gagtgtatta tatcaacact tctgtttgca acactgtaca tctcttgcca catcttctg 420
accgcttca agaagcctgc tgagttcacc acagggtgtc ctgggcggg tctmtgagac 480
agtgggtgat ttgatgtcc tcaactctgt ggtgctagggt atggtgtggg tggcatcagc 540
cattgtggac aagaacaagg ccaacagaga gtcactctat gacttttggg agtactatct 600
cccctacctc tactcatgca tctccttctt tggggttctg ctgctcctgg ctgctggaag 660
acctggagga gcagctgtac tgctcagcct ttgaggaggc agccctgacc cgcaggatct 720
gtaacctac ttctctgtgg ctgcttttag acatggagct gctacacaga caggctctgg 780
ctctgcagac acagaggggtc ctgctgggtg tgtggcttcg tagggcttgg gatacctggg 840
tttccccaag gagagtagcc cctgggttcca ggtgcttgtc gacagcctcc catccctgca 900
cagagaagag gcggaaggct tcagcctgkc aacggaacct gggctacccc ctggctatgc 960
tgtgttctgt ggtgtgtgac ggcctgtctg tgctcattgt ggccatccac atcctggagc 1020
tgctcatcga tgaggctgcc atgccccgag gcatgcaggg tacctcctta ggccaggtct 1080
ccttctccaa gctgggctcc ttgtgtgccc tcattcagggt tgtactcatc ttttacctaa 1140
tggtgtcttc agttgtgggc ttctatagct ctccactctt ccggagcctg cggcccagat 1200
ggcacgacac tgccatgacg cagataattg ggaactgtgt ctgtctcctg gtcctaagct 1260
cagcaattcc tgtcttctct cgaacctgtg ggtcactctg ctttgacctg ctgggtgact 1320
ttggacgctt caactggctg ggcaatttct acattgtgtt cctctacaac gcagccttctg 1380
caggcctcac cacactctgt ctggtgaaga ccttcactgc agctgtgagg gcagagctga 1440
tccgggctt tgggctggac agactgccgc tgcccgcttc cgggttcccc caggcatcta 1500
ggaagaccca gcaccagtga cctccagctg ggggtgggaa gaaaaaaact ggacactgcc 1560
atctgctgcc taggcctgga gggaagccca aggtactctg gacctcagga cctggaatct 1620
gagaggggtg gtggcagagg ggagcagagc catctgcact attgcataat ctgagccaga 1680
gtttgggacc aggacctcct gcttttccat acttaactgt ggccctcagca tggggtaggg 1740
ctgggtgact ggggtctagcc cctgatccca aatctgttta cacatcaatc tgccctactg 1800
ctgttctggg ccattcccat agccatgttt acatgatttg atgtgcaata ggggtgggta 1860
ggggcagggg aaggactggg ccaggggcagg ctccgggagat agattgtctc ccttgctctt 1920
ggcccagcag agcctaagca ctgtgctatc ctggaggggc ttgggaccac ctgaaagacc 1980
aaggggatag ggaggaggag gcttcagcca tcagcaataa agttgatccc agggtttgtc 2040
ttgttttttt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100
nna 2103

<210> 41
<211> 2349
<212> DNA
<213> Homo sapiens

<400> 41
tcgaccacag cgtccgtcta tttctgtgtc tgtatgtatg ggatgtgtct gattgtgtgt 60

ctgtgggtgta	tttctgtgtc	tgtatgtatg	gggtgtgtgt	gattgtgtgt	gtgtctgtgt	120
tatctctgta	tgtatggggg	gtgtctgtgt	gtgattgtgt	gtgtgtctgt	tgtatctctg	180
tgtctgtatg	tatgggggtg	gtctgtgtgt	gattgtgtgt	ctgtgtttat	gtgtgtgtgt	240
ctttgtgtga	ttttttgtgt	atatggcaaa	cccagaacag	aacattatca	ctcacctcac	300
ctggcaaaac	aaaaagcttt	cagagagatg	tgtggccgtc	atgatgtctc	agctgcaggc	360
atttttcaaa	gttatgtttg	atgcagctgg	cacctttctt	ctcggggaca	aatgtgatct	420
ttgatcattg	agttgagaat	gagtttactc	aataatttta	aaatgcctcg	gtgtgagagg	480
atgggctgct	agtaaacatt	tgtggcaaat	atgcgttgcc	ttttcctcct	tcaagtgcac	540
ggggcaggct	gcttctctgc	acgtccatct	gtgggtcagt	ctgtcctgcc	acctcccttg	600
ccccagggtc	tagtgccctg	tgtggatcac	ggctgcccc	gggatgctgg	ggaaggattc	660
agcccaacaa	atagattcgt	tgacgtcctc	tcaacgaatg	tgtgctttta	aaaatgtctt	720
taattttgca	ctccattttc	tacaacaggt	gcacatttaa	aacattttat	tttaaacagc	780
tctgtcttct	cagcactgca	gactattttc	cgctcacttg	acaggcagac	gttgggacta	840
tgttcttttc	ctgccaagaa	tccatcagcc	ccccaatgcc	tgcacctoga	gtccagactc	900
cttagcgttg	cgttgggacg	taattccccg	ttagcatccc	cagccccctc	tccccctgtg	960
agccagcctg	ggttcttttc	tttttccctt	cttccctggt	ttccttcttt	cttttttggg	1020
tttgtaataa	gtaatatata	gtcttgtggg	taagaaatca	aaataataaa	gaagatacgg	1080
cttgaggatt	ctgtctctgc	actggctcct	ccgacaggac	ctggtgcgtc	ccccaggagg	1140
gccttgccat	gtgcgagctg	agacacatgt	tcacactttc	ctcctttttt	gtacaaaagg	1200
tagcacatgg	taaatagcgt	ctgcacgctt	ctccccctgc	tctgtcttcg	ggactactcc	1260
acatccatga	agaagcagca	tccttattct	cttgcacagt	gggcagtggc	tacactgtat	1320
tcactcgtct	ctgctgatgg	gcactattgt	agtgggtgaa	ttagtaacct	agtatacatg	1380
cccttcagcg	tgtgtactgg	tgtatggatc	tcacattccc	agaggctgag	tgaagaataa	1440
aaggcttttt	gacagtattc	ctcaggctact	cacacactcg	ctctgtgagc	cacgtgatca	1500
tgggagcggc	tccctgcaga	gggtgtctgt	gggagaggct	gatgtccttg	tcctgatgggt	1560
gtcccaacct	gagaatgggt	agctgtcttc	ccaatgcaga	ctttcatcac	ttgcactctc	1620
tgcgtgagat	tgaactgctt	ctcatttatg	tttttatctt	tgtttgcccc	tatttttagct	1680
gggttgctaa	tcttttaaat	ttttctaggg	gcttttcata	aattagggag	tttagctgta	1740
gtgatgtgag	ttgaaagtat	tttttcaaat	cgtcattgtc	tcgactttcc	tttttttttt	1800
tcttttgcca	cttaaaatat	gtagttgaat	ctcagtattt	gggggcttct	ggattttgag	1860
tcgtagttag	aaaggtcttc	ctattctgag	gttatcacagg	aattcaccca	tttttttccc	1920
aatattttat	ttccatttta	acttttgacc	cacttggatt	gattttgttg	tatggtgtca	1980
agtatgggtc	caagtttgct	tttctgacct	ggggatccca	acatcgttca	tcaaaatctc	2040
catgttgggc	gggcgtgggt	gctcaagcct	gtaatcccag	cactttggga	ggctgagggtc	2100
ggtggatcac	aagggtcagg	gatcgagacc	accctggcta	acacgggtgaa	atccctatctc	2160
tactaaaaat	acaacaaatt	agccgggtgt	ggtggcaggc	acctgtagtc	ccagctactc	2220
aggaggctga	ggcaggagaa	tcgcttgaac	ccgggaagcg	gagggtgcag	tgagctgaga	2280
ttgcgccact	gcactccagc	ttgggtgaca	gagcaagact	ctgtctcaca	aaaaaaaaaa	2340
aaaaagggc						2349

<210> 42

<211> 1559

<212> DNA

<213> Homo sapiens

<400> 42

attcatgcc	aaacataggc	tttcagtgcc	tattacatat	ggctttcagc	tctctctact	60
gagggatgta	ggagtttatt	tctgaggtct	gagcctcttt	tccttttactt	cccttactct	120
ttcttaagcc	ttctttataa	aaactatgca	tggtctattg	ttttccctttt	tgattccctt	180
tctttttatta	tccccagtag	gagtgacttg	taattctcat	atggttagaaa	ggcagggtctc	240
ctggttgaag	aaaagatcca	cccaagcaag	tcagcagttt	aataaaatttt	tgagggggat	300
ctcaaatgtg	ggaaggattg	ttatataaga	caaccaaatg	atgacatgag	acaataaatg	360
ctataggaat	tatggaggaa	taattagcta	tttatcttct	tggttaggga	agagatatata	420
ttagttgtag	aagtaattac	taacttctac	attttttatt	gtggaaatca	aaaatatata	480
tatgaaaaa	aaatgttata	attgacttca	gtgtcccata	aaccagcttc	caacaattac	540
caaatgtgta	ccaatccttt	acacacatgc	acaggtgtcc	ctccagtatc	tgtggggcat	600
tggttcttagg	accacttatg	gataccaaca	tctatggatg	ctcaagtccc	tgatataaaa	660
tgggtggacta	tatgcatata	acccgtgtac	gtcccgtatt	aaccaaatcc	tccttagatc	720

acttacaaca	cgtaacacaa	tctaaatgcc	acgttaacaa	ccgtcatact	gtattaagcg	780
aataaccccc	ccaaaaatgt	acttctctca	gtccagacgc	cttttttttt	tgtgtgtgga	840
atatcccagc	ccccaaagtc	ccttcaaccc	cctggcacat	aggaggctga	ctgcgtgtgt	900
gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	agacacacat	atctctgaaa	960
tgtaaatatt	ctctctttaa	aaaaattatt	atcacagcta	aacaaattac	cagtaattct	1020
cttatccctca	tatacccggt	gttcagattt	tctagattgg	ctcctaattt	ttttacagat	1080
tatttgaatc	tgtattcaatt	catgtactgt	aatgtttgat	aacttaagta	ccctttatag	1140
gttcccttta	ccacttcttt	attaaattcc	ttgtaatttg	ttgtactaaa	tggattgtct	1200
tcctagaatt	tcctgtagtc	tgaattatgt	ggtaattggt	cacatgttcc	agtgtcctct	1260
tatttctctgt	gagttggtag	ttagatctag	aagcttgatt	aaattcagat	tttctctctt	1320
tagatcatca	acttttagatc	atcaacttgg	atcatttgtt	tcattttgct	tttgatatgt	1380
tgttttttag	aattacctct	taaaattttg	atttaatttt	ataatcatgt	aaaatgttta	1440
taaaatttcca	aattcagatc	agcaaaacac	aataaaatct	attcagagaa	ggcaaaaaaa	1500
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1559

<210> 43

<211> 1766

<212> DNA

<213> Homo sapiens

<400> 43

cggcacgagg	agcactgaag	tattcactac	atgaagtata	ttttgcaactg	tggacacaaa	60
ttagaaaaat	tgcaagtagt	ggtatatatt	aattggcatg	cacttatatga	gcagagtcaa	120
tgtgtctcct	tgtagaatat	tctctgatga	tactcactat	tatccccctt	ctgctaagct	180
ttgtttctgt	tctgaagggc	ataaagcatg	gaaactacat	ttttcagact	ccattaccag	240
aaggatatgg	ttggatttca	gcaatgagtg	ggctttgcat	aaaatttggg	agacgaaaga	300
gaagaaaaac	ctggctgctg	cagggttggaa	cactggcaac	aatagatacg	gagtttgcaa	360
gaagctgcta	agcttctctca	ggaaaattat	ttgtttcaat	atttctggca	tatgggatca	420
tcaattattg	ttttcagtg	ttctggctga	aaattgggtc	aattcttcta	tctgagaatt	480
gttcatttct	gtgcttcagg	aaactaagac	catcactggc	agtttttgtt	gagggatcct	540
tgtgcatcca	tttttcttta	aaacagcctt	cctaactttt	actccccag	cctctatggt	600
tgtgtaagtc	tttaattctt	agagttacat	ttctcttact	cagtatatcc	tagtgcggt	660
tctgttttcc	agacccaacc	ctgactgata	tagtctccat	gtgtttcaga	tgggtgggata	720
gtttggatat	ttgtccctgc	ccaaatttca	tgttgaactc	taatccccag	tgtgttaggt	780
ggggcccaat	gggaggtgtt	cggatcatgg	gggcagatcc	ctcacggctt	ggtgctgtct	840
tcgtgatagt	gagttcttgt	aagatctggt	catttttaag	tgtttggcac	ttgtaccatt	900
tcactgtgtc	ttgtctctgc	tttcaccatg	tgaagtgcct	gttccagctt	cactttttcac	960
catgatttga	aacttctctga	ggcctcccta	gaagccaagc	agatgccatc	accagggttc	1020
ctgtaaagca	tgcagaactg	ttagccaatt	atacctcttt	tttttcaaaa	ttaaaaacct	1080
cttttcttta	caaaatggaa	agaataaagg	tatttcttta	tagcaatgca	agaacggcct	1140
aatacagatg	gctctgcca	tagtgagaaa	attgagacgc	tttctctaga	tggcaaaaaa	1200
gttgtaaaaa	taaaaggaaa	ttattaatat	accgtctatt	gtgatcattt	actaagttaa	1260
gcatattatt	aagaagacaa	gcataagttt	aacacaattt	ggcaatgaat	aaaattgaag	1320
gagagagagc	atatgttggc	tttgctctgt	gaaactcaaa	tgaatgggtc	acctgttcta	1380
gcagctcatg	aaaaattctg	cattgtttat	tatgtgtcag	gatcaacctt	aaattcagtt	1440
ataaaaaagt	tgatgattac	aaaaaaaagg	gaagcactaa	gtaatatagg	tacagagagg	1500
gaagagtgtc	aaatagaatt	ttcaatctgt	gtataaggat	acttaagcat	tttttaagga	1560
aagcagaaa	aagcatgaga	aagtctacaa	tgacatctat	gtcaatataa	caagctggat	1620
atttagagaa	gaaactcttg	attaaatact	ttttatgata	tgaacacaca	catataatat	1680
gacatgactg	tgttcatgga	acataaagaa	attcctctga	ccaaagagaa	ctggaaaaaa	1740
aaaaaaaaaa	actcgagggg	gggccc				1766

<210> 44

<211> 2572

<212> DNA

<213> Homo sapiens

<220>
 <221> SITE
 <222> (2527)
 <223> n equals a,t,g, or c

<400> 44

aattcggcac	gagtcctggac	cttcttagmt	tgcttgatat	caggttggtt	ttgtagccat	60
tttgcttcac	agtcacctgg	aatgccggga	gccccgtgct	atccccgatcc	tctccttgta	120
catgggcgca	cttggtgcgt	gcaccacctt	gtgcctgggc	tactacaaga	acattcacga	180
catcatccct	gacagaagtg	gccccggagct	gggggggagat	gcaacaataa	gaaagatgct	240
gagcttctgg	tggccttttg	ctctaattct	ggccacacag	agaatcagtc	ggcctattgt	300
caacctcttt	gtttcccggt	accttggtgg	cagttctgca	gccacagagg	cagtggcgat	360
tttgacagcc	acataccctg	tggtcacatg	ccatacggct	ggttgacgga	aatccgtgct	420
gtgtatcctg	ctttcgacaa	gaataacccc	agcaacaaac	tggtgagcac	gagcaacaca	480
gtcacggcag	cccacatcaa	gaagttcacc	tctgtctgca	tggctctgtc	actcacgctc	540
tgtttctgta	tgttttggac	acccaacgtg	tctgagaaaa	tcttgataga	catcatcgga	600
gtggactttg	cctttgcaga	actctgtgtt	gttcctttgc	ggatcttctc	cttcttccca	660
gttccagtca	cagtgaaggg	gcattctcacc	gggtggttga	tgacactgaa	gaaaaccttc	720
gtccttgccc	ccagctctgt	gctgcggatc	atcgtcctca	tgcacagcct	cgtgggtccta	780
ccctacctgg	gggtgcacgg	tgcgacctgt	ggcgtgggct	ccctcctggc	gggtcttggt	840
ggagaatcca	ccatggtcgc	catcgctgct	tgctatgtct	accggaagca	gaaaaagaag	900
atggagaatg	agtcggccac	ggagggggaa	gactctgcca	tgacagacat	gcctccgaca	960
gaggaggtga	cagacatcgt	ggaaatgaga	gaggagaatg	aataaggcac	gggacgccat	1020
gggcaactga	gggacagtca	gtcaggatga	cacttcggca	tcattctctc	cctctcccat	1080
cgtattttgt	tccctttttt	ttgttttgtt	ttggtaatga	aagaggcctt	gattttaagg	1140
tttcgtgtca	attctctagc	atactgggta	tgctcacact	gacgggggga	cctagtgaat	1200
ggctctttact	gttgctatgt	aaaaacaaac	gaaacaactg	acttcatacc	cctgcctcac	1260
gaaaacccaa	aagacacagc	tgcctcacgg	tgacgttgtt	gtctctctcc	cctggacaat	1320
ctcctcttgg	aaccaaaagg	ctgcagctgt	gccatcgctc	ctcggtcacc	ctgcacagca	1380
ggccacagac	tctcctgtcc	cccttcacgt	ctcttaagaa	tcaacagggt	aaaactcggc	1440
ttcctttgat	ttgcttccca	gtcacatggc	cgtacaaaag	gatggagccc	cgggtggcctc	1500
ttaaatttcc	cttccgccc	ggagttcgaa	accatctact	ccacacatgc	aggaggcggt	1560
tggcacgctg	cagcccgagg	tccccgttca	cactgaggaa	cggagacctg	tgaccacagc	1620
aggctgacag	atggacagaa	tctcccgtag	aaaggtttgg	tttgaaatgc	cccgggggca	1680
gcaaaactgac	atggttgaat	gatagcattt	cactctgcgt	tctcctagat	ctgagcaagc	1740
tgctcagttct	cacccccacc	gtgtatatac	atgagctaac	ttttttaaat	tgtcacaaaa	1800
gcgcattctcc	agattccaga	ccctgcgcga	tgacttttcc	tgaaggcttg	cttttccctc	1860
gcctttcctg	aaggtcgcat	tagagcgagt	cacatggagc	atcctaactt	tgcatttttag	1920
tttttacagt	gaactgaagc	tttaagcttc	atccagcatt	ctaattgccag	gttgctgttag	1980
ggtaactttt	gaagtagata	tattacctgg	ttctgtctatc	cttagtcata	actctgcggt	2040
acaggtaatt	gagaatgtac	tacggtactt	ccctcccaca	ccatacgata	aagcaagaca	2100
ttttataacg	ataccagagt	cactatgtgg	tcctccctga	aataacgcat	tcgaaatcca	2160
tgcagtgcag	tatatttttc	taagtttttg	aaagcagggt	ttttccttta	aaaaaattat	2220
agacacgggt	cactaaattg	atttagtcag	aattcctaga	ctgaaagaac	ctaaacaaaa	2280
aaatatttta	aagatataaa	tatatgctgt	atatgttatg	taatttattt	taggctataa	2340
tacatttctt	attttcgcat	tttcaataaa	atgtctctaa	tacaatacgg	tgattgcttg	2400
tgtgctcaac	atacctgcag	ttgaaacgta	ttgtatcaat	gaacattgta	ccttattggc	2460
agcagtttta	taaagtcctg	catttgcatt	tgaatgtaag	gctcagtaaa	tgacagaact	2520
attttttncat	tatgggtaac	tgggggaata	aatgggggtca	ctgggagtag	gg	2572

<210> 45
 <211> 526
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (66)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (106)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (484)

<223> n equals a,t,g, or c

<400> 45

ctctgacagc	ttcctctttg	gccaaagccct	gcctctgtac	agcctcgagt	ggacagccag	60
aggtcnagac	tggagccccag	agcccaagat	ggagccccag	ctgggncctg	aggctgccgc	120
cctccgccct	ggctggctgg	ccctgctgct	gtgggtctca	gccctgagct	gttctttctc	180
cttgccagct	tcttcccttt	cttctctggt	gccccaaagtc	agaaccagct	acaattttgg	240
aaggactttc	ctcggtcttg	ataaatgcaa	tgctgcctc	gggacatcta	tttgcaagaa	300
gttcttttaa	gaagaaataa	gatctgacaa	ctggctggct	tcccaccttg	ggactgcctc	360
ccgattccct	ttgstttctt	atccttgcaa	attactccar	atgattycca	aaatctggsg	420
sccttgtaga	ratcttttaa	ctggtcagca	awtwtcaaac	gaaatctcca	aacaggaaat	480
cttntgcctc	ctgcatccac	ccccaaagaa	cttgcacatt	gacgtt		526

<210> 46

<211> 1032

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (974)

<223> n equals a,t,g, or c

<400> 46

gtaaaattgt	caaacattta	tttttaaaaa	taaaatgtgg	tgggcgggta	tgggtggagt	60
tattgggtcc	tggctgtcgc	cactgagcat	tgctgaatgt	tgtcatgact	tgtggacttc	120
tcagagctgc	gagcatgcgg	gtgctctttg	tggagatttg	ctctgtgcct	gtaggaaggt	180
gggtgtgtgg	tgtgcattgc	agcaacattg	gtggaacaga	tgtgtgtgcc	cccattgctg	240
catcagggtt	cattgcacag	gagcctcata	cacactacag	aaaatatgct	cttgtaacct	300
caagtccatg	gggaggcatc	cacacagatg	gcagcagatc	aggaaatgct	ctcagccagt	360
cctcaggggc	agtagggcag	ccttcattcg	ggtgcgcttg	gcagccctga	actttatctc	420
ttcttttcgt	tgtatttctt	taatatctta	ttcagcattt	tagtttcttt	agttggatga	480
gttgttcagt	gtctttgtac	tgtacagttc	tcctgtgctg	ttggaatgga	cgtcattggc	540
tatgttactc	ttttgaaata	caaattttct	tttagaagtt	ttggaaaatc	tgtagtagga	600
ggagtctga	gaatatgaga	agacttcttg	ccacatagtt	gtgcctgccc	gatgtatttt	660
ttcaaaaaaa	ttgaagggtg	acctgttgga	agctcagttc	tgagttaa	tgactacttt	720
gcagagacac	aacactctct	ttstacatg	ttcccttgga	gaaaaataggc	tgggtgcttg	780
gtgggtgatca	gaactggcag	ttattctagg	cctaaaacag	tttggaata	aaataagaaa	840
aaggggccac	atgtggtggc	tcattgctta	atccccacac	tttgagaggc	tgaggcaggg	900
ggattgcgtg	aagccagcct	gggcaacata	aaatttaaaa	attagccaga	catggtggca	960
tgtgcctgtg	ggtncagcaa	cctgggtggc	taagaccgaa	ggatcacttg	agcccaggaa	1020
gtcaaggctg	ta					1032

<210> 47

<211> 2680

<212> DNA

<213> Homo sapiens

<400> 47

ggcacgagtc	ttacagcaag	tacttaacaa	acttagatct	ccataatcag	atctatctca	60
aaccgcaaaa	aaaagctgat	tttattatgt	taaactgata	ggagatcttt	atgagagatg	120
aaagtaagcg	attttaactt	tttaattttt	tttaattttt	ctctgtttct	cacacttgag	180
gcttttctga	aattcaccaa	aagagttttg	gcagtgtgtg	ggaatttacc	agaaccccc	240
attattaaaa	caattggttt	cttatattaa	aatcatcaga	cagggtttgt	gtacacagac	300
atctatgtta	agccactgga	tagaactaaa	ctgtgtgtag	gaacctgttt	tctctaagaa	360
aggactcctt	tgagtctctg	aggatatgtg	cttaaagggt	ttctgtgtca	agctatttta	420
tgaaataaaa	ttttacaaac	aatgaagctt	cattttcaca	cgttgtatct	gtttaatgca	480
gacttggcct	atacagagtt	ttgaaatcaa	gggttttttc	cgtaaaacta	gcttaggagt	540
ttagttagtc	actgtgtctc	tcacctctgt	aggatgagca	cagagatcac	atctgagaag	600
cagggtgtgg	tggtttctgc	tggtcacgtc	tcctctgctt	ctctctggc	tcagcagtta	660
cttttctca	gtcatggggg	ctgcattcag	ctgcgatgcc	agcattaggg	ccaagctctc	720
ccgtctctgt	gacgttttat	tacacgtact	ctttatcagt	agtagctgcc	attccctgag	780
tgctccttcc	accgtgcact	gagaagggtt	gaggatttcc	gaacatcaac	ctcctctaac	840
cctctgtatg	ggaattagat	ttttatcatc	ctgactgttg	gggtcttcta	gagctcaggc	900
aacttgcccc	tggtggggat	ggagcctgcc	tggtggggagt	gtgaatgagg	cttctgttgt	960
ctgtggcttg	gtctctctgt	gttggaact	tcagaagcag	tagattccag	gtgcaacatt	1020
tggtgccccg	acaagtgtga	aaacctgacc	tggtggaggg	ccttcactgg	cagccagtgt	1080
ccaccctga	gcccgggtgag	cacagacgtt	agagccaggg	ccggcggtct	cagagggttg	1140
agccggggac	cagtgggtct	agaggtcgga	gcccggggacc	agtgggtcca	gaggtcggag	1200
ccatgtccct	gcccgtgaagc	tctgaggggt	gggtgggtgaca	tcggggcgat	tgtggcatcg	1260
cgattgtggt	gtcttgggtg	cctttggccg	gcgtgtccac	tgccctgggt	tcattgacgta	1320
ggcaccacat	ggctctgtgt	tgctctcact	acgggtccac	ttctgggtct	ttcttacttg	1380
ctcacccttg	agataaactt	ctatttcat	ttcatttctg	tttatatgtg	ggttttcttc	1440
caggctctga	gttaagccta	taataattgca	atgtgatgtt	ttgaagttaa	ggtgtaatag	1500
agccagtga	ccaagggttc	acacccaggt	gaaatacaaa	tattcagaat	tgagccactg	1560
tggtgccata	ctgattatgt	aatgtgtgat	taacaagtat	aatgtgtcac	tttcaacatc	1620
agtttcatgc	caaagtgtga	ttttattaga	ttatttggga	gttcactttg	ggcccaaagg	1680
ctcgtgtcta	cataataata	acttatgatt	tttctttttg	ttttgtttct	attttttgtt	1740
ttgtgttttt	tgctttctag	accatgccag	agtaattcca	gctttcttta	gttactggat	1800
cacacataatc	cttcctgaga	agagcagtga	ctaaaatgga	atatctcttt	aagaacagct	1860
cctctttaac	aaaaaaactt	aaaagacaaa	tgtgagatgg	gcttagactt	agttctctgg	1920
gaacttgaaa	gacatttatg	ccatattatt	tattcacgtg	tttgttcttg	gtgggcaaga	1980
tgccatctga	ggcttcagat	gagaaattgg	ggtaaaatgg	aaatttttca	cttatttgca	2040
attatatata	ctttgaatta	ctacataaaa	cttgattctg	tttctctact	tattgtaaaa	2100
attgaaaatg	gacattctgt	taagttaaat	gtatagtttg	aagctcataat	atttttatga	2160
agttttgaat	caccttgat	ctgaaaagct	ctgctttaa	aatgctttct	gggtattaaa	2220
atgttctagt	tttaagtagt	tgaatatagt	tgagtctttt	ttctctctct	tactttgtga	2280
atcatatcag	gtacctgttt	ttcctgtttc	gattttcttt	tctgtgatag	aagcagtcgt	2340
cagttcttgg	tattactaag	tggtaaaagc	atcagtcagg	ccgggtgcgg	tggtctcacgc	2400
ctgtaatcct	ggcacttttg	gagaccgagg	caggcggtac	acaacgtcag	gagatcgaga	2460
gcatcctggc	taacacggtg	aaaccccatc	tctattaaaa	atacaaaaaa	ttagccgggc	2520
gtggtggcgg	gcgcctgtag	tcccagctac	tctggaggct	gaggcaagag	aatggcatga	2580
acctggggag	cagagcttgc	agttagctga	gatcgacca	ctgcactgca	gcctgggcga	2640
cagagcgaga	ctccagctca	aaataaaaaa	aaaaaaaaaa			2680

<210> 48

<211> 1730

<212> DNA

<213> Homo sapiens

<400> 48

cccacgcgtc	cgggggcctg	tggtggagtc	caagaggaac	taggcccgtt	tcgggatggg	60
gtggagcgct	gtccttgcct	cctgttgctt	cctcgggggt	ctggggctgg	tccaggtgtt	120
gggggcccac	gccgtggggc	cctggacggc	ttcagcgtgt	ttgggggcag	ctcaggtctc	180
gccctgcagg	ccctgcaagg	agagctctct	gaggttatct	tcagcttcag	ctccctcaat	240

gactcactga	atgagctcca	gaccactgtg	gagggccagg	gcgctgatct	ggctgacctg	300
ggggcaacca	aggaccgtat	catttctgag	at-aacaggc	tgcagcagga	ggccacagag	360
catgctacag	agagtgaaga	gcgcttccga	ggcctagagg	agggacaagc	acaggccggc	420
cagtgcacca	gcttagaggg	gcgattgggc	cgtcttgagg	gtgtctgtga	acggttggac	480
actgtggctg	gggaactgca	gggcctgcgc	gagggccttt	ccagacacgt	ggctgggctc	540
tgggctgggc	tccgggaaac	caacaccacc	agccagatgc	aggcagccct	gctggagaag	600
ctggctgggg	gacaggcggg	cctgggcagg	cggctgggtg	cccttaacag	ctccctgcag	660
ctcctggagg	accgtctgca	ccagctcagc	ctgaaggacc	tactggggcc	tgcaggagag	720
gctgggcccc	cagggcctcc	tgggctgcag	ggacccccag	gccctgctgg	acctccagga	780
tcaccaggca	aggacgggca	agaggggccc	atcggggccac	caggctcctca	aggtgaacag	840
ggagtggagg	gggcaccagc	agcccctgtg	ccccaaagtgg	cattttcagc	tgctctgagt	900
ttgccccggg	ctgaaccagg	cacggttccc	ttcgacagag	tccctgctcaa	tgatggaggc	960
tattatgata	cagagacagg	cgtgttcaca	gtcgccactg	gctggacgct	acttgctgag	1020
cgcggtgctg	actgggcacc	ggcacgagaa	agtggaggcc	gtgctgtggc	cgctccaacc	1080
agggcgtggc	cggcgtagac	tccgggtggc	tacgagcctg	agggcctgga	gaataagccg	1140
gtggccgaga	gccagcccag	cccgggcacc	ctgggcgtct	tcagccctcat	cctgcccgctg	1200
caggccgggg	acacgggtctg	cgtcgacctg	gtcatggggc	agctggcgca	ctcggaggag	1260
ccgctcacca	ctttcagcgg	ggccctgctc	tatggggacc	cagagcttga	acacgcgtag	1320
actgggggtc	cgcccgacgt	gtctacgtcg	ctgaaagaga	cagcgggggc	ggcgggctcc	1380
tggggtctcg	cctgagacgg	ggcacctagc	cctgggcgag	cgccgcaccc	gggcccgcag	1440
cggcacccgg	cccagagcgg	cctctcccca	cgcccggggc	gcgccggctc	agggaggctc	1500
ggggccgccc	atcgagactt	tggccctggc	gcgatccccc	aagaacccct	ccagggccgg	1560
cctgcccagg	agccgatcct	cgcacccctc	gctccctcca	ctggccctcc	aggtcgattc	1620
cctgggctcc	aggtctcccc	gcgcggggcg	cgccaccgcg	catactaaac	gatcgaggaa	1680
taaagacact	tggtttttct	aaaaaaaaaa	aaaaaaaaaa	aaggggccgg		1730

<210> 49

<211> 1275

<212> DNA

<213> Homo sapiens

<400> 49

ggcacgagcc	agctcctctt	ccaagatgtc	ctcccttaaac	tttgccagtg	aaatggagaa	60
gtactgtttg	gggaataata	tgctttcaag	gttttgctta	tttttgataa	tgctcctgca	120
catecttctt	tttttggtta	tcttcattca	acgtcacact	gttgctcttc	tctccaagca	180
ccaccctttt	gttccaacta	atggatcaaa	gagttatagc	agcttttaag	gcctgattac	240
tacctgagga	ggacctttgc	ctaggctatt	gctgaaagata	agaaagatcc	tttttaacaa	300
cactggtgca	attctggaag	gattacaaca	tctatggcag	catcaacaac	cttgggttaag	360
ttgggggtga	agtcaccaag	gagtgtgtaa	atgctatctg	ggaagagata	ttgaagacat	420
ttgtccatga	cttcaaaaaca	ttgctaagga	tgaggatggt	gcaaaaaaaa	aatattcaca	480
aggctttagt	tgaatggca	aacaacttta	agctggatgt	ggaggaggat	gatattgagg	540
aggtcacaga	tgtggtcact	gaggaattga	ctaattgacga	gttactggaa	ccagaacaga	600
aatgcatagc	aagaggaaag	gaaactacag	gagaaaaaaa	gaaaagcccc	caaggatatt	660
cacagtgatg	ggttttaggca	gaagcttttg	cagacctcaa	caaactcctt	aaaacgtttg	720
aaaatgttga	ctccaacatg	aaatgatttt	cattagtaga	gaggaatgtt	catcgtgcac	780
tatctattta	caagcaaatc	tatgatgaat	aaaagaaaca	aaccaaacaa	accaccatga	840
acacattttt	gaaaagagtg	acatctcaag	aagagcctca	ggcagggtcct	tcaggaggta	900
tcccagaaga	agaaggcat	gttatcacag	gacatgacag	ctccctgggt	gttattgccc	960
ttaaaaggcct	tccagttagga	taagggtgtg	agatgggaagg	tggcgatttt	tatgatcctg	1020
acccctacata	ggtctaggct	aatgggttaa	ggtgtgtttg	tgctcttagtt	tttaagaaa	1080
aatgtaaaaa	gtaaaaaata	aaaataacat	taatataggt	aaaagtttat	agaataagga	1140
tataaggaga	gaaaaaaatt	atcttccctc	gtatttgtgt	ttcaagctgt	gttactacaa	1200
aagaataaaa	aagttaagag	aaattgtttt	taaagtaaaa	tagttacaat	aatctaaaaa	1260
aaaaaaaaaa	aaaaa					1275

<210> 50

<211> 1762

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (447)

<223> n equals a,t,g, or c

<400> 50

gggtggcctag	agatgtctgt	gccgcggttg	cagttgtctg	gcacgcctct	gcccgcacgc	60
ccgctccacc	gccgtagcgc	ccgagtgctg	gggggcgcac	ccgagtcggg	ccatgaggcc	120
gggaaccgcg	ctacaggccg	tgctgctggc	cgtgctgctg	gtggggctgc	gggcgcgcac	180
gggtgcctg	ctgagtgggc	agccagctct	ccggggaggg	acacagaggc	cttggtataa	240
agtcattttac	ttccatgata	cttctcgaag	actgaacttt	gaggaagcca	aagaagcctg	300
caggagggat	ggaggccagc	tagtcagcat	cgagtcgtga	gatgaacaga	aactgataga	360
aaakttcatt	gaaaacctct	tgccatctga	tggtgacttc	tggattgggc	tcaggaggcg	420
tgaggagaaa	caaagcaata	gcacagctct	ccaggacctt	tatgcttgga	ctgatggcag	480
catatcacaa	tttaggaact	ggtatgtgga	tgagccgtcc	tgccgcagcg	aggtctgcgt	540
ggtcatgtac	catcagccat	cggcacccgc	tggcatcgga	ggccctaca	tggtccagt	600
gaatgatgac	cggtgcaaca	tgaagaacaa	tttcatctgc	aaatattctg	atgagaaacc	660
agcagttcct	tctagagaag	ctgaagggtga	ggaaacagag	ctgacaacac	ctgtacttcc	720
agaagaaaaca	caggaagaag	atgccaaaaa	aacattttaa	gaaagtagag	aagctgcctt	780
gaatctggcc	tacatcctaa	tccccagcat	lccccctctc	ctctctcttg	tggtcaccac	840
agttgtatgt	tgggtttgga	tctgtagaaa	aagaaaacgg	gagcagccag	accctagcac	900
aaagaagcaa	caacacctct	ggccctctcc	tcaccaggga	aacagcccgg	acctagaggt	960
ctacaatgtc	ataagaaaac	aaagcgaagc	tgacttagct	gagacccggc	cagacctgaa	1020
gaatatattca	ttccgagtgt	gttcgggaga	agccactccc	gatgacatgt	cttggtgacta	1080
tgacaacatg	gctgtgaacc	catcagaaag	tgggtttgtg	actctgggtga	gcgtggagag	1140
tggattttgtg	accaatgaca	tttatgagtt	ctccccagac	caaattgggga	ggagtaaggga	1200
gtctggatgg	gtggaaaatg	aaatatatgg	ttattaggac	atataaaaaa	ctgaaactga	1260
caacaatgga	aaagaaatga	taagcaaaat	ctctttatct	tctataaggga	aaatacacag	1320
aagggtctatg	aacaagctta	gatcaggctc	tgtggatgag	catgtgggtc	ccacgacctc	1380
ctgttggacc	cccacgtttt	ggctgtatcc	tttatccag	ccagtcaccc	agctcgacct	1440
tatgagaagg	taccttgccc	aggtctggca	catagtagag	tctcaataaa	tgctacttgg	1500
ttggtttgat	ctaactttta	agggacagag	ctttacctgg	cagtgaataa	gatgggctgt	1560
ggagcttgga	aaaccacctc	tgttttctct	gctctataca	gcagcacata	ttatcataca	1620
gacagaaaaa	ccagaatctt	ttcaaaagccc	acatatggta	gcacaggytg	gcctgtgcac	1680
cggcaattct	catatctgtt	ttttcaaaag	aataaaatca	aataaagagc	aggaaaaaaa	1740
aaaaaaaaaa	aaaaaactcg	ag				1762

<210> 51

<211> 2059

<212> DNA

<213> Homo sapiens

<400> 51

cggcacgagg	tttgggtctaa	gagtgagcag	aactgtgttg	agatgggagc	acattttggtt	60
gtgttcaaca	cagaagcaga	gcagtggaag	gaggcatgca	tccccagctg	ataccttcgg	120
ttattgctgt	agttttcatc	ttacttctcg	gtgtctgttt	tattgcaagt	tggttggtga	180
ctcatcaca	cttttcacgc	tgttaagagag	gcacaggagt	gcacaagtta	gagcaccatg	240
caaagctcaa	atgcatcaaa	gagaaatcag	aactgaaaaag	ttgctgaagg	gagcacctgg	300
aactgttgtc	ctattgactg	gagaaccttc	cagttccaac	tgctattttc	ctctttactg	360
acaacaagac	gtgggctgag	agttgaaagg	aactgttcag	ggatgggggc	ccatctgatg	420
accatcagca	cggaaagtga	gcagaacttt	attattcagt	ttctggatag	acggctttcc	480
tatttccttg	gacttagaga	tgagaatgcc	aaaggtcagt	ggcgttgggt	gggaccagac	540
gccatttaac	ccacgccaga	gtattctggc	ataagaatga	acccgacaac	tytcaggggag	600
aaaactgtgt	tgttctgggc	ttataaccaa	gataaatggg	cctggaatga	tggttccttgt	660
aactttgaag	caagtaggat	ttgtaaaaata	cctggaacaa	cattgaacta	gaaactcaga	720

aagtggctct	tgtgatggaa	agagaaaaga	aaaaccaatt	agaataaggc	agaatgtacg	780
tgcgtcattg	gaacacagaa	aacatgctgg	ttcatacagc	gttttttagtc	ataatggctc	840
tttttatttt	gtttgattca	ttcgagacaa	catgtgtgta	tgtgtgtgtg	tgtgtgtgta	900
gataatgtgg	tttttgtatg	gtgtttgatg	gaaggaataa	tctttctttg	ctttcttagt	960
agtatttcaa	gggtgtttact	tttcaattgg	tgtgcactga	atgcatgtat	ggaagaatag	1020
cgtgaataat	gcaatctctt	tgtcattttt	tcccctttct	cagactctta	gctcttaaaa	1080
ttcaaaagat	gggatattct	aactggtagt	ggtgcatcat	ttttaaccca	aattattgcaa	1140
gcactttaaa	gatttgaaac	cacattttta	ttgtttgatg	tttcattttc	agacttttta	1200
atgtcagtc	ttacaattac	attgcatgag	gaaaattttt	ccagaacaac	agtgtggaat	1260
agttctgaat	tatgctgttc	tacagatgga	aaaaagtcca	aatgccttta	aaaatttact	1320
tcttactcca	cccaacacgt	ttttgcaaa	caagaagtct	ttgtaagaca	ccttaaacaa	1380
agtccttcaa	ttctacagca	gaggaaataa	aatccccag	agcccaaagg	gctcaccttc	1440
acattgttag	ttcatgacag	acccagggtg	gcttcattag	agataacata	cattcccttt	1500
gggtatcacag	gaagttactg	gggattactc	gacctcatta	cttagctaac	gactggataa	1560
aattttctaa	ttgtttgaag	taacattgta	ttcgtgtttg	cattattaat	ttgaatagaa	1620
aataatcaca	ttttcaaccc	atttatacaa	attgttaatg	tttctttaga	gctgtataac	1680
tatagtttga	actagcaagg	aagttattgt	tttgacaacc	agaaattatg	cttttctggt	1740
gcatgaaaca	ctaattgcaa	agggcagtc	catccaactt	taataaaaata	tggtggctct	1800
tcttaaaatt	ttcaatttgc	taatttttcc	tggaacccta	actttcctac	tactactttt	1860
aaaagcagta	tttattttgt	aaattagcta	ctgatttttt	gggtttgaaa	tactgggtgt	1920
tgattgtgtg	gtgggtgggtg	tggtgatgga	tgcgtgtagc	tgtgttaaaa	gctgatactt	1980
ctagctgaaa	tggtgctaaa	taaatttgac	tgtgtttgcc	atattcacta	aaaaaaaaaa	2040
aaaaaaacta	gagggggggg					2059

<210> 52

<211> 3282

<212> DNA

<213> Homo sapiens

<400> 52

cccacgcgtc	cgacttaaaa	gagaagcttt	agctgccaaa	gattgggaaa	gggaaaggac	60
aaaaaagacc	cctgggctac	acggcgtagg	tgcagggttt	cctactgctg	ttcttttatg	120
ctgggagctg	tggtgttaac	caactaggaa	ataacgtatg	cagcagctat	ggctgtcaga	180
gagttgtgct	tcccaagaca	aaggcaagtc	ctgtttcttt	ttcttttttg	gggagtgtcc	240
ttggcagggt	ctgggttttg	acgttattcg	gtgactgagg	aaacagagaa	aggatccttt	300
gtggtcaatc	tggcaaagga	tctgggacta	gcagaggggg	agctggctgc	aaggggaacc	360
agggctggtt	cggatgataa	caaacaatac	ctgctcctgg	attcacatac	cgggaatttg	420
ctcacaaatg	agaaactgga	ccgagagaag	ctgtgtggcc	ctaaagagcc	ctgtatgctg	480
tattttccaaa	tttaaatgga	tgatcccttt	cagatttacc	gggtgagct	gagagtcagg	540
gatataaatg	atcacgcgcc	agtatttcag	gacaaagaaa	cagtcttaaa	aatatcagaa	600
aatacagctg	aagggacagc	atttagacta	gaaagagcac	aggatccaga	tggaggacct	660
aacggtatcc	aaaactacac	gatcagcccc	aactcttttt	tccataattaa	tatttagtggc	720
ggatgatgaag	gcatgatata	tccagagcta	gtgttgga	aagcactgga	tggggaggag	780
cagggagagc	tcagcttaac	cctcacagcg	ctggatgggtg	ggctctccac	caggtctggg	840
acctctactg	tacgcacgt	tgtcttgga	gtcaatgaca	atgccccaca	gtttgcccag	900
gctctgtatg	agacccaggc	tccagaaaa	agccccattg	ggttccttat	tgttaaggta	960
tgggcagaag	atgtagactc	tggagtcaac	gcggaagtat	cctatttcatt	ttttgatgcc	1020
tcagaaaaata	ttcgaacaac	ctttcaaatc	aatccttttt	ctggggaaat	ctttctcaga	1080
gaattgcttg	attatgagtt	agtaaatctt	tacaaaaata	atatacaggc	aatggacggt	1140
ggaggccttt	ctgcaagatg	tagggtttta	gtggaagtat	tggaacacaa	tgacaatccc	1200
cctgaactga	tcgtatcac	attttccaac	tctgttgctg	agaattctcc	tgagacgccg	1260
ctggctgttt	ttaagattaa	tgacagagac	tctggagaaa	atggaagat	ggtttgctac	1320
attcaagaga	atctgccatt	cctactaaaa	ccttctgtgg	agaattttta	catcctaatt	1380
acagaaggcg	cgtctggacag	agagatcaga	gccgagtaca	acatcactat	caccgtcact	1440
gacttgggga	caccagggct	gaaaaccgag	cacaacataa	ccgtgctggg	ctccgacgtc	1500
aataacaacg	ccccgcctt	cacccaaacc	tcctacaccc	tgttcgtccg	cgagaacaac	1560
agccccgcc	tgcacatcgg	cagcgtcagc	gccacagaca	gagactcggg	caccaacgcc	1620
caggtcacct	actcgtgt	gccgccccaa	gaccgcacc	tgccccctgc	ctccctggtc	1680

tccatcaacg	cggacaacgg	ccacctgttc	gccctcaggt	cgctggacta	cgaggccctg	1740
caggctttcg	agttccgcgt	gggcgccaca	gaccgcggct	ccccgcgct	gaacagcgag	1800
gcgctgggtg	cgctgtctgg	tgctggacgc	caacgacaa	tcgcccttcg	tgctgtaccc	1860
gctgcagaac	ggctccgcgc	cctgcaccga	gctgggtccc	cgggcggccg	agccgggcta	1920
cctggtgacc	aaggtggtgg	cggtggacgg	cgactcgggc	cagaacgcct	ggctgtcgta	1980
ccagctgctc	aaggccacgg	agcccgggct	gttcgggtgtg	tgggcgcaca	atggggaggt	2040
gcgcaccgcc	aggctgtctga	gcgagcgcga	cgcagccaag	cacaggctgg	tggtgcttgt	2100
caaggacaat	ggcgagcctc	ctcgctcggc	caccgccacg	ctgcacgtgc	tcctgggtgga	2160
cggcttctcc	cagccctacc	tgctctctcc	ggaggcggcc	cgggccagg	cccaggccga	2220
cttgctcacc	gtctacctgg	tggtggcggt	ggcctcgggt	tcttcgctct	tcctcctctc	2280
ggtgctcctg	ttcgtggcgg	tgcggtctgt	caggaggagc	agggcgccct	cgggtgggtcg	2340
ctgctcggtg	cccgaggggt	cttttccagg	gacgtgaggg	gcgctgagac		2400
cctgtcccg	agctaccagt	atgaggtgtg	tctgacggga	ggccccggga	ccagtgaatt	2460
caagttcttg	aaaccagtta	tttcggatat	tcaggcacag	ggccttgga	ggaagggtga	2520
agaaaattcc	accttccgaa	atagcttttg	atttaatat	cagtaaagtc	tgtttttagt	2580
ttcatatact	tttggtgtgt	tacatagcca	tgtttctatt	agtttacttt	taaattctcaa	2640
atttaagtta	ttatgcaact	tcaagcatta	ttttcaagta	gtataccctt	gtggtttttac	2700
aatgtttcat	catttttttg	cartaataac	aactgggttt	aatttaatga	gtattttttt	2760
ctaaatgata	gtgttaaggt	tttaattctt	tccaactgcc	caaggaatta	attactatta	2820
tatctcatta	cagaaatctg	aggttttgat	tcatttcaga	gcttgcatct	catgattcta	2880
atcacttctg	tctatagtgt	acttgctcta	tttaagaagg	catatctaca	tttccaaact	2940
cattctaaca	ttctatatat	tcgtgtttga	aaaccatgtc	atttatttct	acatcatgta	3000
tttaaaaaga	aatattttct	tactactatg	ctcatgacaa	aatgaaacaa	agcatattgt	3060
gagcaatact	gaacatcaat	aataccctta	gtttatatac	ttattatttt	atctttaagc	3120
atgctacttt	tacttggcca	atattttctt	atgttaactt	ttgctgatgt	ataaaacaga	3180
ctatgcctta	taattgaaat	aaaattataa	tctgcctgaa	aatgaataaa	aataaaacat	3240
tttgaaattt	gtgaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aa		3282

<210> 53

<211> 1860

<212> DNA

<213> Homo sapiens

<400> 53

aattcggcac	gagcagcagg	gagaagagga	tgatgagcat	gccaggcccc	tggccgagtc	60
cctgctcctg	gccattgtctg	acctgtctct	ctgcccggac	ttcacgggtc	agagccaccg	120
gaggagcact	gtggactcgg	cagaggacgt	ccactccctg	gacagctgtg	aatacatctg	180
ggaggctggt	gtgggcttcg	ctcactcccc	ccagcctaac	tacatccacg	atatgaaccg	240
gatggagctg	ctgaaactgc	tgtgtacatg	cttctccgag	gccatgtacc	tgccccagc	300
tccggaagt	ggcagcacca	acccatgggt	tcagttcttt	tgttccacgg	agaacagaca	360
tgccctgccc	ctcttcacct	ccctcctcaa	caccgtgtgt	gcctatgacc	ctgtgggcta	420
cgggatcccc	tacaaccacc	tgctcttctc	tgactaccgg	gaacccctgg	tggaggaggc	480
tgcccagggtg	ctcattgtca	ctttggacca	cgacagtgcc	agcagtgcc	gccccactgt	540
ggacggcacc	accactggca	ccgccatgga	tgatgcgat	cctccaggcc	ctgagaacct	600
gtttgtgaac	tacctgtccc	gcatccatcg	tgaggaggac	ttccagttca	tcctcaaggg	660
tatagcccg	ctgctgtcca	acccctgtct	ccagacctac	ctgcctaact	ccaccaagaa	720
agatccagtt	ccaccaggag	ctgctagtct	tcttctggaa	gctctgcgac	ttcaacaaga	780
aattcctctt	cttcgtgtctg	aagagcagcg	acgtcctaga	catccttgte	cccacctctt	840
tcttctctca	cgatgcccgg	gccgatcagt	gtaagaccag	ggtgggggtg	ggatgctggg	900
ggcttctcctg	gcccgcagggt	ggagggtctg	ctatctgccc	tgactcccag	gagctcagcc	960
tggcactctg	acgtcctagg	agggaggcct	ccaaaatggt	cactgcctgg	gcccctcctt	1020
ccccctccca	cccgagtgc	ccctcccacg	ccacctgccg	cagctcgggt	gggcctgatg	1080
cacattgggtg	tcttcatctt	gctgcttctg	agcggggagc	ggaacttcgg	ggtgcggctg	1140
aacaaacctt	actcaatccg	cgtgcccacg	gacatcccag	tcttcacagg	gaccacgcc	1200
gacctgtctca	ttgtggtgtt	ccacaagatc	atcaccagcg	ggcaccagcg	gttgcagccc	1260
ctcttcgact	gcctgtctac	catcgtgggt	aacggtaggg	gcccggagct	tgactcccc	1320
gcgctcacca	ccaggtggtg	tcagccacag	gccaacctgc	cacttgctcc	aggtttcaag	1380
ccccctccagc	ccacgaaag	aacgcaagct	tgtgtacctt	tcatagtagc	cacattaaaa	1440

aaaaaggggtg	acaacaattg	aataaatatgt	tttatctatc	atcatctcaa	catgtagcca	1500
atatgaaaat	tactaatgag	atattttaata	ttcccttctc	acactcagtc	ttgtcgccca	1560
ggctggagtg	cagtggcacc	atctcagctc	actgcaacct	ccgccccctg	ggttcaagca	1620
attctcatgc	ctcagcctcc	tgagtaggtg	ggattacagg	cgtgtgccac	cacacctggc	1680
taatttttgt	atcttttagta	gagacggggt	tttgccatgt	tgtccaggct	ggctctgaac	1740
tcctgacctc	agatgatatg	cccacctcag	cctcccaaaag	tgctgggatt	acaggcatga	1800
gccactgcat	ctggcctttt	taaaaaaaaa	aaaaaaaaaaa	actcgagggg	gggcccggta	1860

<210> 54

<211> 770

<212> DNA

<213> Homo sapiens

<400> 54

aattcggcac	gagccgaggg	gcgaatccat	acatccgtcg	tgtggactcg	ccctgacatc	60
ccctcaacgc	ttagactcgg	cattcttttc	attgcctgat	gccgtttcat	gccttctttg	120
cgtttctctg	ctctggccct	cctgctggcc	atccttcccg	cacttcccaa	cgcccatgcc	180
gcgcggggga	ttggcgccct	gatcgggcgc	ggctcccagg	cctcagccaa	ggaagagccc	240
cagagcaacg	cgcaaccacg	cgccgatgag	cgcaaaacaac	gactgctcag	ccaggccgag	300
gaaacccggc	agcgcctgac	cgatctcaag	gccgaactcg	ccggcgcacc	caaggagatc	360
agcgaggccc	agcgtacgct	ctccaaaactg	gtgagcgagg	acaacagcga	tctgcccagag	420
cgctcttcca	agcttttcggt	gccggtactg	gagcaacgcc	tggcgggccc	cgtggacgag	480
ctcgccctct	ggcaacaagc	gtcagcgcg	gccaacagca	tgtcatcag	cgcgagacc	540
cgcccgagc	gcgccaggg	cgatatcagc	aagaaccagt	tgcgcacga	cgagatcaac	600
ggcctgctga	aaagcggctg	ggagaacaac	aagccgctga	cggaacgaac	tcgcgcgctg	660
ctcgagagta	cttctagagc	ggccgcgggc	ccatcgatct	tccaccggg	tggggtacca	720
ggtaagtgtg	cccaattcgc	cctatagtga	gtcgtattac	aattcactgg		770

<210> 55

<211> 1093

<212> DNA

<213> Homo sapiens

<400> 55

cagatttcggc	acgaggtttt	ttttaagatt	tatttttttaa	aaattatttt	tgtggacttg	60
ggatcaaatg	atggcaccta	cttttgggaa	tctgtagctg	tgtcttgaga	attgccatcg	120
gtcatgtgtt	gcaccgttct	ctgtatgttt	acgtcttttg	gactggcttc	tcccaggatt	180
ctttctgttt	tttggttttt	tgatttgggc	tttatttttt	tctgtgtact	gtactatatt	240
gtaaaaggga	ttttagcaga	gacttttagt	tttggggcaa	gaggagaaca	ggaatgctgg	300
gctgtttact	ttaggtggag	aaccatctt	cagacctttg	gactattttc	tttcaactgc	360
agtgtataga	aaaaccaaac	tacgacctca	gagcagagta	ttaatgaaaa	gcacaaaaaa	420
aggaactaag	ttcagcgagg	ggtgggggga	ggggggagat	ttttcttttg	aaaaataatg	480
actcttagga	catttgtttt	tcagttcaag	tgtcttccag	cactgtcttg	tctcccaata	540
taccaaccca	ctggcacatt	tttctcttgt	tttctctctc	cgatttttgt	ctgtctcttc	600
agttaagtgt	ttcttctctt	tgtgcccccc	gctggtgacc	ctctgcttcc	ctctctcttt	660
ccctttggca	gctgcaatac	acagtgttat	tttggggaaa	taaattctagc	aaagcctcgc	720
cttccatgac	gagcgtcttc	ttggctctga	gagggaaagg	tctgtctctg	ggatgctctc	780
tgttcttttt	tccccctaag	tcttctctct	tcccatcata	cccttccctg	cccaccttgt	840
tttctgttct	ccttttatta	ggaattccca	agtgaaattt	attaatgtgg	gagtgggaaca	900
gatgctaaaa	gctatccagg	attttgtttc	tgtttgtttt	aaattttgtg	gttcttctcc	960
tttctctccc	ctcccatgcg	taagacgttc	tgtgtaacct	ccattaaaatt	tggtaaaaaa	1020
ccactcgcca	gagctgtggg	gtcagaaaaa	taaaatatac	tgtttctctam	aaaaaaaaaa	1080
aaaaaaaaact	cga					1093

<210> 56

<211> 632

<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (29)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (46)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (94)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (162)
<223> n equals a,t,g, or c

<400> 56
cgcaattaat gtgagtttagc tcactcatna ggcaccccag gctttncact ttatgctttc 60
cggctcgtat gttgtgtgga attgtgacgg atancaattt cacacaggaa acagctatgc 120
catgattacg ccaagctcga aattaaccct cactaaaggg ancaaaaagct ggagctccac 180
cgcggtggcg gccgctctag aactagtggg tcccccgggc tgcaggaatt cggcagcaga 240
ctatgtatat atgtttaata tctgtctttt gaaatgcaga aatagttaa atgtttcttt 300
gtctattttt cttttttttt aatgctaccc agggaaaata tttcatatca tttttaagt 360
gcctgcctca atgtatatatt atttcttttg aaacaaaaag gttctggaaa ctgtttttct 420
gtagctttta atgaatagggt gagcaaaatc tatatgggat gtaatttttt tgttcagtct 480
cttaaaaaat actttgtttt ggtacatttg gttgtgcttg tggggaaaat aaaaacgcag 540
agatccttat atatttatgt taaagtaata ttttattatc tacataaaac agaaatgcac 600
aataaaaaaa aaaaaaaaaa gctcgagggg gg 632

<210> 57
<211> 2687
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (1614)
<223> n equals a,t,g, or c

<400> 57
gtacaccatg ggcctccacc tccgccccta cegtgtgggg ctgctcccgg atggcctcct 60
gttctctttg ctgctgctaa tgcgtctcgc ggaccagcg ctcccggcgg gacgtcaccc 120
cccagtgggt ctgggtccctg gtgatttggg taaccaactg gaagccaagc tggacaagcc 180
gacagtgggt cactacctct gctccaagaa gaccgaaagc tacttcacaa tctggctgaa 240
cctggaactg ctgctgcctg tgcacattg actgctggat tgacaatatc aggctggttt 300
acaacaaaac atccaggggc acccagtttc ctgatgggtg ggatgtacgt gtccctggct 360
ttgggaagac ctctctcactg gagttcctgg accccagcaa aagcagcgtg ggttctctatt 420
tccacaccat ggtggagagc cttgtgggct ggggctacac acggggtgag gatgtccgag 480
gggctccccta tgaactggcg cgagcccaaa atgaaaaagg gccctacttc ctggccctcc 540
gcgagatgat cgaggagatg taccagctgt atggggggccc cgtgggtgctg gttgcccaca 600
gtatgggcaa catgtacacg ctctactttc tgcagcggca gccgcaggcc tgggaargaca 660

agtatatccg	ggccttcgtg	tcactgggtg	cgccctgggg	gggcgtggcc	aagaccctgc	720
gcgtcctggc	ttcaggagac	aacaaccgga	tcccagtcac	cgggcccctg	aagatccggg	780
agcagcagcg	gtcagctgtc	tccaccagct	ggctgctgcc	ctacaactac	acatgggtcac	840
ctgagaaggt	gttcgtgcag	acaccacaaa	tcaactacac	actgcgggac	taccgcaagt	900
tcttcaggga	catcggcttt	gaagatggct	ggctcatgcg	gcaggacaca	gaagggtcgg	960
tggaaagccac	gatgccacct	ggcgtgcagc	tgcactgcct	ctatgggtact	ggcgtcccca	1020
caccagactc	cttctactat	gagagcttcc	ctgaccgtga	ccctaaaatc	tgctttgggtg	1080
acggcgatgg	tactgtgaac	ttgaagagt	ccctgcagtg	ccaggcctgg	cagagccgcc	1140
aggagcacca	agtgttgctg	caggagctgc	caggcagcga	gcacatcgag	atgctggcca	1200
acgccaccac	cctggcctat	ctgaaacgtg	tgtcctcttg	gccctgactc	ctgtgccaca	1260
ggactcctgt	ggctcggccg	tggacctgct	gttggcctct	ggggctgtca	tggcccacgc	1320
gctttgcaaa	gtttgtgact	caccattcaa	ggccccgagt	cttggactgt	gaagcatctg	1380
ccatggggaa	gtgctgtttg	ttatcccttc	tctgtggcag	tgaagaagga	agaaatgaga	1440
gtctagactc	aagggacact	ggatggcaag	aatgctgctg	atggtggaac	tgctgtracc	1500
ttaggactgg	ctccacaggg	tggactggct	gggcctcgtg	cccagtcctc	gcctggggcc	1560
atgtgtcccc	cctattcctg	tgggcttttc	atacttgctc	actggggcct	ggcncsgcag	1620
ccttcctatg	agggatgtta	ctgggctgtg	gtcctgtacc	cagaggtccc	agggatcggc	1680
tcctggcccc	tcgggtgacc	cttcccacac	accagccaca	gataggcctg	ccactgggtca	1740
tgggtagcta	gagctgctgg	cttccctgtg	gcttagctgg	tggccagcct	gactggcttc	1800
ctgggcgagc	ctagtagctc	ctgcaggcag	gggcagtttg	ttgcgttctt	cgtgggtccc	1860
aggccctggg	acatctcact	ccactcctac	ctcccttacc	accaggagca	ttcaagctct	1920
ggattgggca	gcagatgtgc	cccagtcctc	gcagctgtgt	tccagggggc	ctgatttctc	1980
cggatgtgct	attggcccca	ggactgaagc	tgcctccctt	caccctggga	ctgtgggtcc	2040
aaggatgaga	gcaggggttg	gagccatggc	cttctgggaa	cctatggaga	aagggaatcc	2100
aaggaagcag	ccaaggctgc	tcgcagcttc	cctgagctgc	acctcttgct	aaccccacca	2160
tcacactgcc	acctgcccct	aggtctctac	tagtaccaa	tggtcagca	cagggtgag	2220
gatggggctc	ctatccaccc	tggccagcac	ccagcttagt	gctgggacta	gcccagaaac	2280
ttgaatggga	ccctgagaga	gccaggggtc	ccctgaggcc	ccccaggggg	ctttctgtct	2340
gccccagggt	gtcccatgga	tctccctgtg	gcagcaggca	tggagagtca	gggctgcctt	2400
catggcagta	ggctctaagt	gggtgactgg	ccacaggccg	agaaaagggg	acagcctcta	2460
gggtgggttc	ccaaagacgc	cttcasgctg	gactgagctg	ctctcccaca	gggtttctgt	2520
gcagctggat	ttctctctgt	gcatacatgc	ctggcatctg	tctccccttg	ttcctgagtg	2580
gccccacatg	gggtctctgag	caggctgtat	ctggattctg	gcaataaaa	tactctggat	2640
gctgtaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaag	ggcggcc		2687

<210> 58

<211> 619

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (526)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (619)

<223> n equals a,t,g, or c

<400> 58

tcacagctt	tcacctggat	tatcgtgaca	gcctcctact	gcttctctat	catgtggcca	60
gagctatctt	ccataaatgc	attgcatagt	tgatcaagtc	actctctggc	ctaaaacctt	120
ccttggtctc	ctgtgcccct	caggataaag	tctggacccc	tcagcatggc	ttgtgagact	180
catgggtgtc	ttgtccctgc	tcacctctct	ggctctcatca	cttgccttct	tgcattcttg	240
gtcccagcct	cctgtatcca	gagatgcagt	ggctctccat	tgccactctg	attcctctct	300
tcttttgggc	acagagaaa	ggtactttct	ctgtcaaatc	tcaacttaga	cttgacttcc	360
tccaaggagc	tttggctata	ctctctctct	ccgaccccca	ccctggcata	ctacacagat	420

cactctgggc	tcacttgcc	gcctaattgg	catctcccca	gtagactgta	agctccttga	480
gggcaaggat	tgtgttgga	ttttgtatt	aacagtgcct	ggcttngtgc	ctggcaccta	540
gaaagcactc	aataaatggt	tgtttaatga	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	600
aaaaaaaaa	aaaaaaaaa					619

<210> 59
 <211> 1378
 <212> DNA
 <213> Homo sapiens

<400> 59	
tcgacttctg	ggatgtagat tctcctttat tgaattcttc attttcactc gcagtatgat 60
ttattttttt	cacactgtat ttacactgc atatgtttct cttataacga gggtcacttt 120
cttctgccc	aggattggaa tcaattacag ctatgcaatg tttttgttt agcatcttct 180
taattacagg	cttggcagaa gagttttgtg aacaactctc catttcactg gctgaagaag 240
aaatacaact	gtcctctaca gttgaacatt tctgtatgac agcattttca tggtaatgag 300
gattaggttc	atactttggt ttttccaaac cggggaaaca aacaacagcc aaaaaccagt 360
gtgcagcttc	attaaggggt acaaaaataa aatccttctc aaaaatatct acgtgccggg 420
tccatgtttt	tactctccca tgcggttttt gctgtattga cagattagtt gtttcattgat 480
ttctcctctc	tctctgatta aggcgtttat agaaaaaaga actgaatata tgaattcggg 540
cagcgtcttc	cttcttcagt ttttcaagca ccaagtattt caaataaaag tctataataa 600
catcatttaa	aaattctcct tcatttagac agtgcaggtc ctccattggta acagagatgc 660
ctcccttagc	tggagggtgt ggatatacta tcaacttttc tactggggcca atgaagatgg 720
tgtggttttc	tccagtttct tcttcttcat caaaaaactg aaattcttgt ttgcttctaa 780
gttgtatttt	agattcaaat gatacagttt taatttttgt ttctttttgc ccacaacttc 840
ctttgatgct	ctcttcatag gttcttgtac aggcaacaag tctgccatta gcttcttcaa 900
aggggaattt	cgcaaaaaaa ttggagatgt tattctttat accaatttca ttaatgatac 960
tttcaaatac	catatttgcg ggaggatcaa ggccattttg aaaaattaaa attatatatt 1020
gttcttctaa	atttgtaaat ttatttactc ctttacaatc attccaaact ttatcctcct 1080
tattcatttg	cagttggatg ctcagctttt gataaactgc tgggaattgct tgaagaaaca 1140
ctacaggtaa	ttttcggaca ttacaccatt cacatttagt tagatcagag gtatttaata 1200
taatctctac	aggatcatgg tctggttcgt ctagctgtat cttgataaaa tctaaacaaa 1260
aaattacagg	ctctattaac agccggaaga gtgttcttac tcgtcgacgc ggccgcgaat 1320
tcccgggtcg	acgagctcac tagtcggcgg ccgctctaga ggatccctcg agggggccc 1378

<210> 60
 <211> 1126
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (21)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (35)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (49)
 <223> n equals a,t,g, or c

<220>
 <221> SITE

<222> (99)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1012)

<223> n equals a,t,g, or c

<400> 60

tccggcttcg taatgtggtg ngaaattgta acggnataac aatttcacnc aggaaccagc	60
tatgcccattg attacgccaa gtctgaaatt aaccctcant aaaggaccaa aagtgaagtt	120
ccaccgcggt gccggcccgcc tctagaaact agtggatccc ccgggctgca ggaattcggc	180
acgagccggc ccgtaccgcc aggcgatcgc gctgatggcg gcgctggcag cagcggccaa	240
gaaggtgtgg agcgcgcgcc ggctgctggt gctgctgttc acgccgctcg cgctgctgcc	300
ggtggtcttc gccctccgc ccaaggaagg ccgctgcttg tttgtcatcc tgctcatggc	360
ggtgtactgg tgcacggagg ccctgccgct ctcagtgcag gcgctgctgc ccatcgctct	420
gttccctctc atgggcatct tgccctccaa caaggctctgc cccagctact tcctcgacac	480
caacttcctc ttcctcagtg ggctgatcat ggccagcgcc attgaggagt ggaacctgca	540
ccggcgaatc gccctcaaga tccctgatgtc tgttggagtc cagccggcca ggctcactct	600
ggggatgatg gtgaccacct cgttcttctc catgtggctg agcaacaccg cctccactgc	660
catgatgctt cccattgccca atgccatcct gaaaagtctc tttggccaga aggaggttcg	720
aaaggacccc agccaggaga gtgaagagaa cacaggaata gaaccacaata ctttctcttc	780
tgaggaaagg ctgaaacttc aagctccctc tgtgataaga cttggtcaga taactgagtc	840
tgttcaatgg aatatgagtg gaaatgatgt gtgcaacttc cgggttctgt ccttctgccc	900
gggtggaatg tgaatatgat ggcacctggg acccaaagac aggagccaca tcttgagaga	960
tagatggcag atctgcccct gtggctttgg atcatttacc tcagtgaaca cnacaagcat	1020
tatccatgaa accatagggt ttgtgtgcta gttctagtct ttaaaatatg aattaaatta	1080
aatacgtatc tgttaaaact taaaaaaaaa aaaaaaaaaa ctcgag	1126

<210> 61

<211> 2078

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (337)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (492)

<223> n equals a,t,g, or c

<400> 61

gattattttc aatagattgt caataaatta tcaaattatg gaaatatcct ttagattcct	60
gtgataatgc tttgtgcatg ttttctatta tctagaaaat gaaaacagca gtcatttgac	120
cagccatggt atgacaccag ataataattct gcaaggatct tccaaacagg gagaaactgc	180
aatgggaaca atctttggat atctacattg tgtcaaatgt tatgtgcttt atttcatttt	240
catacttata acagctgtat atcacagttt ttattatcca cattatagag gcaaggcact	300
gatatacagg acttgagtaa ttggcttaca cagcaantgt agttgaaata gaagccagga	360
ctgtctgatt cccaagacag tgttcttgta ctgcaggtaa gcctgggtgt gctgggkttc	420
cctagcactg agtattgcag aaatacggag atttattttac tcagggtgaat gatgaagcag	480
aatctttatt tngcttttta aattatkgtat gggkckkgag ttaagctagt tcaccaatga	540
gtctcaagg gtgaagaact ggcctgtgta gttttaaaac aaagcagcag ttttcttcat	600
ttgtagatat tatatttaat ttaccttgca tttccaatag cattaaatat tagaagcgta	660
taagcttttt ctttttatgc ctaggggata tggcttaaca ctgagaaaca cttgagaaac	720
actgccacat tctaaggaaac aaccaataat aagaaaaatt atttcatata aaaggatgca	780

gcttgcatat	aaaatgctga	attcaactga	agttcattat	gctgccttta	tgtgagctca	840
ggggtgacct	tcatgctctg	gactcttaaa	ggatccagtg	tcggatacag	ctctgccatt	900
ggagcaatgt	gttccataac	taggggtgcc	taagtatatc	ccttcttctg	ktatctatag	960
tttttcttta	aagggtgtgct	tttcaaaaaca	tgatgttatt	tcatatacmt	atatatatat	1020
atgcacacat	acatatacmt	catgcatgka	tcatatatgt	gtgtgtatgt	atgtatctat	1080
cttcaatctt	tcaggaaaagt	ttgtaaaawa	tagccctaag	tagaattatg	gggaaacagt	1140
tgcagaactg	ctgaaaattgg	ctcagatgaa	aaggaaaagag	caaaaatcta	ggttattcac	1200
aggaagttag	tccagctgtg	ttttgcgagg	catgaaccag	gctgaatcag	agcaagaaca	1260
ggttgcaatc	ttaggtcaca	tggcaattaa	attccagcta	tagaatattg	aagacatgta	1320
agtgtctaat	ttaaggatta	ttcaatgcag	tgaataaatt	tatatcattc	ccaaaagtat	1380
gcacatcaga	taatgagtta	actgcatgta	acttgaaaac	cttaaataatt	taactgaaag	1440
gatttttgtg	atagaaaagat	gtcttagtca	atgtaaaaaat	gccagtgatt	tgtgtagcat	1500
gtccttttgc	aaaagtccatt	atataaactg	aagtcactta	tctttgtttc	caatgtgata	1560
atattccaca	tttaacagggt	gataattatt	gaaaatatta	gtttgcctca	ctaataattgt	1620
gtccgagata	ttctagttta	aaccaagtaa	ctaatactgt	aaaagaagta	ttccttttat	1680
gaagccagca	ttatcctgat	actagaactg	ggaagagaca	cacacaaaaa	gaaaacttca	1740
ggccaatatc	gctgatgaac	atcaatgtga	aaatcttcaa	taaaatactg	gcaaaccgaa	1800
tccagcagca	tgaggcggtg	ggatcacaa	gtcaggagat	cgagaccatc	ctggctaaca	1860
tggtgaaacc	ccatctctac	taaaaatata	aaaaattagc	cgggcatggt	gatgggcacc	1920
tgtagtccca	gtactggggg	aggctgaggc	aggagaatgg	cgtgaacctg	ggaggcgga	1980
cttacagtga	gccgagattg	caccactgca	ctccagcctg	ggtgacagag	caagactctg	2040
tctcaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aactcgta			2078

<210> 62

<211> 762

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (10)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (12)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (42)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (219)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (747)

<223> n equals a,t,g, or c

<400> 62

cgagtcagtn	ancgagaaag	cggaagagcg	cccaatcgca	anccgcctct	ccccgcgcgt	60
tgccccgattc	attaatgcag	ctggcacgac	aggtttcccg	actgaaaagc	ggccagtgag	120
cgcaacgcaa	ttaatgtgag	ttagctcact	cattaggcac	cccaggcttt	acactttatg	180
cttccggctc	gtatgttgtg	tggaaattgtg	agcggatanc	aatttcacac	aggaaacagc	240

tatgaccatg attacgccaa gctcgaaatt aaccctcact aaagggamca aaagctggag	300
ctccaccgcg gtggcgcccg ctctagaact agtggatccc ccgggctgca ggaattcggc	360
acgagcacia atgcctatac aaaacttttt tccatgtcct caatactttg taccagatta	420
tttatatttc tcatggtttg gtttcttttt cttgtcttca tttttctgaa ggtaaaagg	480
gatttttttc ccccttttct gatctgtaac ttgttttgta ttggatgat tacaggcgtg	540
agccaccgcc tccagccaca aatacttttt agtcgccata agcataatca agaaatcatt	600
ttgcaaatgg taagcttttc atgttggtga ttctttccca tgattagaga ggtaaaaagt	660
gscctgggat gcataaagat gtcgtagacm tmtgcaatct gccacattat ttctttaaaa	720
aaaaaaaara aaaactcgag ggggggnacg gaaccaggg cg	762

<210> 63

<211> 1094

<212> DNA

<213> Homo sapiens

<400> 63

tcggcacgag gtcaatcaag tgaaatatca tgtaaactgt ccagcagctt tgaaagtaga	60
gaatgaacaa ggcccccttc ccacccaccc tgtggaaagc ccgtctgggt tgggtgctctc	120
ctggacagcg tcttgccggt cactcttggc catctcccgg tgcgtgggtc agatgtgggt	180
cctgcttttc tgcccccttc ctctctgtg cctgcctgcc tctgctgtgc cgggccagt	240
ccttggtggc cagtggagt gacaccagct gcgactgcgt gggaggggct ggcattgccg	300
ctgccactgc agggcttggg cggctgacat gggacgaggc ttgcacagct gccagctcct	360
gtctcgctga ctttttttat acagttttgt ctggggccacc gccttcagtg ccacgggcc	420
cttgccgttc aggtgcttc tcatagatga acaaggccct gccccgtgtc cttacccttt	480
agagctgttt aaattcaa atgactgaaac tgaatatgaa aaatccaggc cctcagccgc	540
ccaggccatg tttcaagtgc tccatggcca catgtggctg gtggacagtg cagctctaga	600
acattccatc accacagagg gttctgctgg acagtggcct tgggggctgt tttgagggtc	660
cgcctgtcag tctcctggca tcaaagtcac tctgccattg tcaagttaca gttattttcc	720
ttttacctcc aagccactat gtgcgtatgc tgctatgtgt ctgtatttcc cgctaaactt	780
cctgtcacgg aggggaaggt gccacaggcc cagctcctgg aggggggttg gatgtctggg	840
tgggggggaa ggtgccacag gccacgctcc cggagggggg ttggatgtct ggggtggggc	900
cagggcgcca cagttccagc tcccggaggg ggtttggatg tctggggcgg gctttcttca	960
tgttccatgt atgataacgg tgactggggg gtttacagag agaggcatat aaatagtgtt	1020
ggagtgttgt tttcgcta ataaatgttt gaatagctaa aaaaaaaaaa aaaaaactcg	1080
agggggggcc cggt	1094

<210> 64

<211> 1361

<212> DNA

<213> Homo sapiens

<400> 64

ccggggtcga cccacgcgtc cgaatttctt tagtgatatt tatgtacttt tgtattaatt	60
acattgggta gactgttcca taaatgttga atagatccta tttgttgggt aatgctattg	120
ttgtgtatcc tgctgatttt ctgtgtagt ggtctatcag ttgttgggag aagagtgttg	180
aagtctacaa ctataattgt gtatttgtct attactcctt tcagttcttt cagttccatt	240
tcacatattt tccagctctt gatttggtgca cattaatgat tactatgtct ttttggtgaa	300
tagaccattt tataatttta tgatgtacct ctcttgctac tttctttgct ctgaagtcta	360
cctgatatta atatagtcat tctagctttc tttggagtaa tggtagtgtg atatatcttt	420
tttctgctct ttaactttta cctatcttat atcattatat tattttaagt gaggtttttg	480
tagatagcac ttaacttgtt ctgtcatgtt tttaaatcca ctgtacaagt acaaatctgt	540
ctttattttt cattttttat cttattttga gacagagtct cactttgaca ctcagcctag	600
aatgcagtgg tctgatctcg gctcactgca acctctgctt ccgaggggtc agcaatcctt	660
tcacctcagc ctcccaagta gctgggacta caggcatgaa ccaccatgcc tggctaattt	720
tttttttttt ttttttttga gagatgggga tttaccatgt tgctgaggct ggccctgaac	780
ccttgggctc aaacgacctg cccagctcag cctcccagag aaatatgatt ttaatttagca	840
catttagatc atttactttc aatggaatta ttgatgtatt acaattcaag tctgtcatct	900

tattttattgc	cttccatttt	tccctccgtt	tttcatttct	gtgctttttt	tctttgcatg	960
taaacatggt	ttaggattct	gttttgatgt	atttagagta	tttttttagtg	tatcggttta	1020
tgtagctttt	ctagtggatc	ctttaagtat	tacattatat	atacataact	aagaagtata	1080
tcagtgtcat	cattttacca	gttcaagtaa	agtacagaaa	tctttgtccc	ttttacctgt	1140
cctgcttata	atgtaaagt	cttaaagatt	tccctgacac	acatttagaa	ccataatcaca	1200
gtgttggttt	tgcttcagtt	gtgaaatata	atttagaaac	tcatgagaat	aaaaacccat	1260
tatattttatt	tataaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	agggcgggccg	c		1361

<210> 65
 <211> 947
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (67)
 <223> n equals a,t,g, or c

<400> 65	
ctttctatag	60
acgcgtncgg	120
ctgaaggctc	180
tccttacaca	240
gtctccctcg	300
cctcagtaact	360
tcagctctcc	420
tttaccctcc	480
ggcatagggg	540
attagatcca	600
tgggtcacca	660
tcccgagact	720
tccacagatt	780
gtccccccca	840
ctacagctgc	900
aggatccaag	947

<210> 66
 <211> 1376
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (18)
 <223> n equals a,t,g, or c

<400> 66	
ggcggggcggc	60
gcgcggggcgg	120
cggcgctgcc	180
atagtacagg	240
ggcatgcatg	300
ataaaaagaag	360
acttccatcc	420
atcaacaatg	480
catttcggcc	540

aaagagtctg	ggccccctgg	ccacagtgcg	aggggtgggtca	ccgtctcctc	tgccacccat	600
tacgtcgctg	agctgaacat	ggatgacctt	cagagcagtg	cctgctactc	acccccacgca	660
gcctacgccc	agagcaagct	ggcccttgtc	ctgttcacct	accacctcca	gcggctgctg	720
gcggctgagg	gaagccacgt	gaccgccaac	gtgggtggacc	ccgggggtgg	caacacggac	780
stctacaagc	acgtgttctg	ggccacccgt	ctggcggaaga	agcttctcgg	ctggttgctt	840
ttcaagaccc	ccgatgaagg	agcgtggact	tccatctacg	cagcagtcac	cccagagctg	900
gaaggagtgt	gtggccgtta	cctatacaac	gagaaagaga	ccaagtccct	ccacgtcacc	960
tacaaccaga	aactgcagca	gcagctgtgg	tctaagagtt	gtgagatgac	tggggctcctt	1020
gatgtgaccc	tgtgatatcc	tgtctcagga	tagctgctgc	cccaagaaac	acattgcacc	1080
tgccaatagc	ttgtgggtct	gtgaagactg	cgggtgtttga	gtttctcaca	cccacctsc	1140
cacagggctc	tgtcctctag	ttttgagaca	gctgcctcaa	cctctgcaga	acttcaagaa	1200
gccaaataaa	cattttggag	gataatcacc	ccaagtgggtc	ttcaaccata	aactttgtga	1260
ttccaaagtg	cccagttgtc	acaggtgcca	taaataatta	cattttccaa	cataaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaagggc	ggccgc	1376

<210> 67

<211> 2434

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (10)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (12)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (27)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (73)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (75)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (103)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (130)

<223> n equals a,t,g, or c

<400> 67

ctgggggtan	tncaagaacc	ctctgtngga	cttagatgtc	aagctctttc	ctttgggcag	60
cgtgttccct	tttnccaggt	agtggtcgtg	gtaaactaaa	ttingccggt	cgctttccat	120

ttcctgacan	ttgagatgga	atgccttgac	cattgggtgct	ctgacagaga	agtcattggag	180
tcattgccat	ttcctgggtg	cccttttgga	atgtgatcct	gttagtagag	gttttctagc	240
ttctactaag	atattttctt	ccctaaccat	catacacttg	gcatgtttca	ttcccatctc	300
ctttcccttc	accttaaagg	agactacccc	tttgcccgat	attgtcaacc	taattttctc	360
tcgtactctc	tctagtgaat	gatgtgctac	caagtatatg	ccaggctgtg	agaggattat	420
actgagtagt	agaaagaagc	taatttgaaa	taaaaattat	ttgtataatt	aagaaagcag	480
attagatgca	catgggtcaac	aggaagtga	ctgtatgtct	gctagttaga	ttcaaacat	540
cataaagatg	atagcatgtc	aatatattag	cctagccatt	atgttagcct	ttgttaggtg	600
ggcagctttt	ctgctttttc	ccttcctctg	tgggtgacaac	ggaggaaata	tccaacagaa	660
atacgtctaa	cagggaaatt	gggatcatag	tttatatgca	tctgatttga	aaggagtatt	720
gaggaagggt	ttcatatag	attcatcttt	ggattaaaaa	gaacatttat	gaaatcaagc	780
cttctaacac	tagttataat	tgagaagcaa	cagtaactcc	gtggacagca	atcaagctta	840
aaattgtaaa	taaatatggg	gataattcag	ttgttgcaaa	aaaagggcag	aattcagtag	900
aataaagtcc	ttttctctta	caggtattaa	atgaggacag	agaacctcag	gtgttcttat	960
gctagtgtct	gctgagtgc	tactaagaaa	gcaattccaa	atagatgtat	acatctagag	1020
agagtgggtat	tagagattca	gtgtatgtat	ttattcacat	gagaggaaac	tggaaataaa	1080
ttccataaat	tattggaata	taatcccata	aattatcacc	ttttatgact	ggaaaatatt	1140
tgccaatgaa	gaaatgggtc	gtaggatatt	gtcttaagat	ttttggctgt	ttaataaaaa	1200
tgtaacttta	acgggtttct	atagtgcct	ttataaagt	tattgtctaa	aatatttttg	1260
tatcatgtgc	ctttgaaatt	tgacagctga	ttgggtgtgt	ggatttctgc	ccagccattt	1320
atcagtatta	tcattttatt	cagttagctg	caggtgtatt	agacaaacga	gacttaggta	1380
aggaatggaa	cctttcctgt	ggtttgactg	cacatcacac	cagaagactc	cagtatccct	1440
cattccagaa	tgaggaaaaa	gtattctaca	aagaacctaa	tcacctctgt	gaaatctatg	1500
ggatggaaac	agtgtggcct	taggagtcac	atagtctctg	catgggtggg	aggatcatga	1560
tggaatatgt	gaatttctac	ttctagaagt	tgtgaaatag	gtcctgcact	tttgagaat	1620
gtccttcttt	aaacctggct	tattccacag	ctgtagctga	taacatgacc	tggggcttag	1680
ctgctctagc	cctgggttct	tggagacctc	acactgcctg	gccccgtggc	atccacctaa	1740
ggactgcctg	ctttctgggtc	acatgtggac	cctgatacga	ctaagcgggt	acatatgtgg	1800
ttgtgcaaaa	gctttctgtt	taatgcatag	tgttaccgat	ttacatcttg	gttttcagtg	1860
gcactatgtc	taggaggcaa	tatcctttta	aacagtgtct	tggctaagat	agatacttgt	1920
gaatcaaaga	tagcacagaa	atgaactaag	tatatcccat	ttggaattat	attttgatac	1980
tatttaaaat	ggtttcacct	gttaaagggc	caacagaact	cctgggttta	cttttgtaat	2040
tactgtacag	aaaatttcaa	gagtgtttga	gtgcttgta	tcagggtgtt	tccttaataa	2100
gtagggatat	gatcatttac	aggaattata	tatgaaaaaa	gtttttgaaa	tgtatttttg	2160
tgatgtgcta	tgttgagggg	aaaccaaata	tttatgattt	taaaacattc	gtatgaaaac	2220
attgtacaat	gtaatatgct	caactttctc	aattttttgc	taatttttct	aagatacatt	2280
aaaaatgttt	tatatttttt	tttaagtaaa	atggacctcg	taagaaaaat	aaaaatacca	2340
gaacataaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2400
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaa			2434

<210> 68

<211> 1086

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (10)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (77)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1056)

<223> n equals a,t,g, or c

<400> 68

ttgaaacacn cttttgagca ttaggtccca gctccaccgc ggtggcgggc gctctagaac	60
tagtggatcc cccgggntgc akkaattcgg caccagcaac ctgtagttag ttggctgtca	120
ctcagcagct ctggaaaatt acctgtctgc actgaatttc ctcatcagta aaatggaaat	180
gattatagta ctgaccttgt aggattatc taaaaatcag agaagtccat gcagcttaga	240
acagtgccag gcacatggca aatgctatgg cataacttaag catcttcctc tgggtgect	300
cgtcatcacc atgtgattgt gccctgcttg tccctgtttc acttttcaga gggagaaaag	360
tggccaactt taagaatcaa aattctgatg ttacttcggg aaatgcatag agccagagag	420
acacaatttg acttagtatg atccacatca tcccctcagg ctgaatagtg gtggcatgca	480
catctatacc aaaatgtttt accttttttg tagaaggaaa atatttgtat cttctattcc	540
atatcttaga tctttataag agcacttaag ttcaacctcc taagaaactg ccaattttgt	600
tgatcatgat agtctgcaca gattttcgta ctatttagtg ktgggagtgc cttagggacc	660
atcaacaaca ggscctcctc tttatccatg agactactga ggccttggag gttatttgtc	720
catccatggc ggtgacacrg cttagaggtta tctggttagt agcccaacta agattagaac	780
ccaggaattt tgatttaatt tcaaatgacc tcttttattt ctgcatccgg gaaggagaag	840
aggaaagtaa acgggaaatt catgttatct atggaaaagt tatcagtttg atgtttatta	900
aatgatttgt ttcaggagat tgtttacaac ttttatcttc ctagacaata attttctgta	960
agagtaaagt catggtcatt aaggtagtca tattaatgta ttcagtaasc tgtgaagaaa	1020
aatatatatc aatgttttcc aataaaatac agtgantacc tgaaaaaaaa aaaaaaaaaa	1080
aaaaaa	1086

<210> 69

<211> 1262

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (568)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (639)

<223> n equals a,t,g, or c

<400> 69

cctyctatag gkmaagctgg tackcctgca ggtaccgggc cggaattccc gggctgaccc	60
acgcgtccgc ttacaacttc tctyttgttt ctgtgtttct tccacaccat tatggcgat	120
ttactgatct acaaccgggt ttaagttagt tgcctgctta aaattttctc atgaattctc	180
attacatttt caaacaaaat ccaaagtttt taccatgacc ttgtggatat cttcttgacc	240
tcatttacta ctgttatctt ccttattcat tccattccac ccacagtggc cttcttgcaa	300
gtcagagaat gcaacaaata tgctcctgcc ttggggcatt tgcacttctg tctctttggc	360
ctggtcactt caccctccaca tctctcatct tttatgactt tcttccctata tttggatctc	420
tgttcaaagt tcacccctca aaaaggcctt caaaactacc ctatttaaaa tagcacctgc	480
cactgccatc tccataccca ttttcatttt tccatagcac ttatccctac ttggcattat	540
gttatatat tttttatttg ktactgnt gtctccccc ctggactgta agctccatga	600
ctgtaagtac atagatgaat tacaaaaatga atcaatggng aawtatcctt gmcataat	660
atttaatgat tttgmctcca ttttaagtaa aaaaaaata ccattttttt cactcttcaa	720
agtgatatag tttaatcttc taaactacat tttctctatt tcccgattta attaaatcag	780
tgatataaaa aaaagagtga tggggatctg tgaaagaaga ctaaaataga tgccaggaaa	840
tactaaaact gctgaagttt agtggatat tttttcttta cacacagtat tatttgagtt	900
actaattgtg tcaactgaatt acagaataag caaaatacta ggtaaacaga atcacgcttg	960
ggggctatat tttgtgtaaa atttgtgta tgcaaaaata atattaaata ttttaattact	1020
acagttttgt tatttctttc ttattttagg aaatgatttg cagctgagtg aatcaggaag	1080
tgacagtgat gactgaagaa atatytagct ataaataaaa atttatacag catgtataat	1140

ttattttgta ttaacaataa aaatttcctaa gactgaggga aatatgtctt aacttttcat	1200
gataaaagaa attaaatttg attcagaaat ttcaaaaaaa aaaaaaaaaa aagggcgggc	1260
gc	1262

<210> 70
<211> 1642
<212> DNA
<213> Homo sapiens

<400> 70	
ggcgcgctgc ccggartcgc ctgggttgcg ctgccggcca cgtccccgcg cggggcctca	60
ggctccttcc tactgtccga gggccaccag gccgcccggg gcctgctgcg cccggatgcg	120
tctgttacta gagtggagag tctaccttcg tctcacatgt gccacaaagg atggcatggc	180
ccgggagtgcc cccaccacgt ggctttcacc ccctgcaaag ccagacttcg cccagcgaca	240
cagtgtcaag cccacagctc tccaaggagg aagatggtcc aggctgggag catcycctta	300
gcagcagcct ctgatccctt ggccaagcag gagggaacca ttagcagcct gaggagctgg	360
ctggctggga gcctcgggga ccgcccagcc ttgctcccag ctcaccacca agatgtggac	420
agctcttggt ctcatttgga ttttytcctt gtccttatct gaaagccatg cggcatccaa	480
cgatccacgc aactttgtcc ctaacaaaat gtggaaggga ttagtcaaga ggaatgcac	540
tgtggaaaca gttgataata aaacgtctga ggatgtaacc atggcagcag cttctcctgt	600
cacattgacc aaagggattc ggcagcccam ctcaactcta tggaaagtcac aacagaggac	660
acaagcagga cagatgtgag tgaaccagca acttcaggag gtgcagctga tgggtgtgac	720
tccattgtct ccacggctgt ggctccagc acgactgcgg cctccattac gactgcggcc	780
tccagtatga ctgtggcctc cagtgtctcc acgactgcag cctccagtac aactgtggcc	840
tccattgtct ccacgactac agcctccagt atgactgcgg cctccagcac tcccatgaca	900
cttgactctc ccgcgcccac gtccacttgc acagggcgga ccccgctccac taccgccact	960
gggcatccat ctctcagcac agccctcgca caagtgccaa agagcagcgc gttgccaaaga	1020
acagcaaccc tggccacatt ggccacacgt gctcagactg tagcgaccac agcaaacaca	1080
agcagcccca tgagcactcg tccaagtctt tccaagcaca tgcccagtga caccgcggca	1140
agccctgtac cccctatgck tccccagca caaggctcca ttagccaggt gtcagtggac	1200
cagcctgtgg ttaacacaac awataaatcc acamccatgc cctcaaacac aaccmcwgag	1260
ccccctaccc aggcctgtgt agacaaaact ctccctcttg tgggtgctgtt actcgggggtg	1320
acccttttca tcacagtctt ggtttttgtt gccctgcagg cctatgagag ctacaagaag	1380
aaggactaca cccaggtgga ctacttaatc aacgggatgt atgcggactc agaaatgtga	1440
ggggggcggg ggccctggcg gaggcctggc ccttctctcg tcccttcctt ttgcccctga	1500
gaccaaacca agtgcctcca aattcttttg gtgcaattga ggagatatgc cagatgctta	1560
aacacattta attgctgtca gattaattcc atgatcacta aagagttgct gcttttttca	1620
taaaaaaaaa aaaaaaaaaa gg	1642

<210> 71
<211> 921
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (4)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (9)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (11)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (15)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (20)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (901)

<223> n equals a,t,g, or c

<400> 71

gcgnggggna	naggnaagcn	ccccactatt	gggttcacaaa	gctggagctc	caccgcggtg	60
gcggccgctc	tagaactagt	ggatcccccg	ggctgcagga	attcggcacg	aggtctgagc	120
agataagatt	aagggctggg	tctgtgctca	attaactcct	gtgggcacgg	gggctgggaa	180
gagcaaagtc	agcgggtgct	acagtcagca	ccatgctggg	cctgccgtgg	aagggaggtc	240
tgtcctgggc	gctgctgctg	cttctcttag	gctcccagat	cctgctgata	tatgcctggc	300
atttccacga	gcaaaggac	tgtgatgaac	acaatgtcat	ggctcgttac	ctccctgcca	360
cagtggagtt	tgctgtccac	acattcaacc	aacagagcaa	ggactactat	gcctacagac	420
tggggcacat	cttgaattcc	tggaaggagc	aggtggagtc	caagactgta	ttctcaatgg	480
agctactgct	ggggagaact	aggtgtggga	aatttgaaga	cgacattgac	aactgccatt	540
tccaagaaaag	cacagagctg	aacaatactt	tcacctgctt	cttcaccatc	agcaccaggc	600
cctggatgac	tcagttcagc	ctcctgaaca	agacctgctt	ggagggattc	cactgagtga	660
aaccactca	caggcttgtc	catgtgctgc	tcccacattc	cgtggacatc	agcactactc	720
tyctgaggac	tcttcagtgg	ctgagcagct	ttggacttgt	ttgttatcct	atcttgcatg	780
tgtttgagat	ctcagatcag	tgtttttaga	aatccacaca	tcttgagcct	aatcatgtag	840
tgtagatcat	taaacatcag	cattttaaga	aaaaaaaaaa	aaaaaaarct	cgaggggggg	900
nccggtaccc	agggcggaag	a				921

<210> 72

<211> 906

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (34)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (833)

<223> n equals a,t,g, or c

<400> 72

ggaaattctc	cctcactaat	tggaacaaaa	gctngagctc	caccgcggtg	gcggccgctc	60
tagaactagt	ggatcccccg	ggctgcagga	attcggcacg	aggggaagaga	gaggggaggg	120
tgagcagagg	acaggccggg	agttttcccg	gaacggagga	agagcagtg	aggtctgccg	180
gatgaggctg	ctgtgtggcc	tgtggctgtg	gctctccttg	ctgaaagtcc	tgaggggcca	240
gacccaaccc	cccctgccac	tcccgcctcc	gatgcagagc	ttccaaggaa	accagttcca	300
gggggaatgg	ttcgtcctgg	gcctggcggg	caacagcttc	aggccggagc	acagggcgct	360
gctgaacgct	ttcaccgcaa	cttttgagct	aagtgatgat	ggccgctttg	aggtgtggaa	420

tgcatgact	cgaggccagc	actgtgacac	atggtcttat	gtgctgatac	cggcagccca	480
gcctgggcag	ttcactgtgg	accacggtgt	gggcaggagc	tggttgctgc	ctcccgggac	540
gctggaccag	ttcatctgcc	tgggcagagc	tcarggcctc	tcggatgaca	acattgtctt	600
cccagatgtg	actggargtg	ccctggacct	carcagcctg	ccctgggtgg	cagccccagc	660
ctgaccactc	agacagccgc	ggccccaag	gcctgactct	tcttggtgga	gggcgaggct	720
ggtcacccca	ggccagcgtc	tgttgaagga	tgaagcagct	cctgtccggc	ccagccctgc	780
ctcacagctg	tgcgagctct	gccctcctca	gctctcaaac	ctgaataaat	gcnccaagcc	840
cagaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	ctcgaggggg	ggcccgttac	ccaattcgga	900
agattg						906

<210> 73
 <211> 680
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (7)
 <223> n equals a,t,g, or c
 <220>
 <221> SITE
 <222> (9)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (15)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (16)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (22)
 <223> n equals a,t,g, or c

<400> 73						
cactcantng	aacannagct	cnagctccac	cgcggtggcg	gccgctctag	aactagtgga	60
tccccgggc	tgcaggaatt	cggcacgaga	tatttcgctg	gacctagaa	aagccaccac	120
gacctgtggg	ccatgatgct	accccaatgg	ctgctgtgc	tgttccttct	cttcttcttt	180
ctcttctctc	tcaccagggg	ctcactttct	ccaacaaaat	acaacctttt	ggagctcaag	240
gagtcttgca	tccggaacca	ggactgcgag	actggctgct	gccaacgtgc	tccagacaat	300
tgcgagtcgc	actgcgcgga	gaaggggtcc	gagggcagtc	tgtgtcaaac	gcaggtgttc	360
tttggccagt	atagagcgtg	tccctgcctg	cggaacctga	cttgtatata	ttcaaagaat	420
gagaaatggc	ttagcatcgc	ctatggccgt	tgtcagaaaa	ttggaaggca	gaagttggct	480
aagaaaatgt	tcttctagt	ctccctcctt	cttgtgtsct	cctcctycty	cacctgctct	540
cctccctacc	cagagctctg	tgktcaccct	gttccccaga	gcctccacca	tgagtggagg	600
gaagtgggga	gtgattgaaa	ttaaagagctt	tttcaatgaa	aaaaaaaaaa	aaaaaaaaaa	660
aaactcgagg	gggggcccgg					680

<210> 74
 <211> 1633
 <212> DNA

<213> Homo sapiens

<400> 74

```

ggcacgagca tcagtaagta ggagctaggg aagagaaagc atgcaaaagg gcaagaatca      60
ggagaattga tatagtagct ggacagaact ctagatgtgt gtgtgtgaga gaaagagagg      120
cagggagaag gagggaggag ttacccccac gatgacctcc aacttccctt tctgcacctt      180
catcctgggg atagcacagg ctccaggcctg cctcgggttg cctggcgatt ggctggcct      240
gggctcaggg gtgggggagg ggctgcacca cattaggacc tgccgtactc caatcccatg      300
cagtcctcct gctcctgctg ctgctgtcct gggtctgtgt catgccaggc ttccttgtgt      360
cctgcgtctg tggccgggtc ctgccaacct gtctagtctt ttcaggcttg aggccctgca      420
ttgctctttc tggtcctctc cctcctctcc cgctcccatc ctagcctttt ttgggttccg      480
ggatctgctg acagactttc ttcttgtctg ctgctgtctt acatttcaga agacccctct      540
ggaactgccc atggctgtgg tccacctgct ggtagcaacg cctgtttacc aaatgctaga      600
taatctgcca ctccccctct cagccgcca aatggctgta gctgccacct gttgtctgtt      660
tccacacctt tgtacactgc attctgctg tcaactctga cagctgccca gcagctgcag      720
ctgaggccgt cctgccagca cctaccgcag ccccatgggt ggctcctgtc agttcctggc      780
gtccccagcc tgctctatct gctgtactta cagagccaca tctgggtgaa tacagactg      840
gggtagctct gatgggcca tggcacactc actccagag aggcgggcat cctctctctg      900
gacctctagg gaagggcagg gtctggagcc caataataaa acacagtaat gataatcata      960
gctaattgtt actgagaact taggatgtga taggctcaat gatttaccag aattactcta      1020
tactgcataa ctaccctatc aggttaagtac tattaatatc ccttttccca catgaagaaa      1080
ctaaggctaa ccaatagaaa tgaacctgtg taatgtccca tgggcataat tcatggagcc      1140
aggattcaga accagggtgt ttacttcagg gctgaagctc ataaccacta gaccaaagcg      1200
cctcctccgt gacctctggg gaaggccag caacatctc tattgcgtcc aatgacttct      1260
ccctggctct agatccactg actcacaggg gtagggttaa gggttaaggcc agagtctcag      1320
ctaagctttg gatactttct tctggatttg gagaaggctg gaataactga atttctgtct      1380
atcttcagga gcggcctact agagccacac atttcccagc tctccttgtg tgtttccag      1440
accctctctc catcgtgtcc tctccttag gctcctggaa agttttcaga gagaatcacc      1500
cagtggtaac attgttaaac aaaacaggaa aatgggactt gtgtgtatat atgattaaat      1560
tattaattga tggatctacc ttcttagctc gtgccgaatt cgatatcaag cttatcgata      1620
ccgtcgacct cga

```

<210> 75

<211> 1022

<212> DNA

<213> Homo sapiens

<400> 75

```

ttccccgggtc gacccacggc tccgcccacg cgtccggctt ggggccagca cctgtctca      60
aagatggcaa aatgaggcta gtcttggatg agctagctgg tgtgggttcc aaccatagga      120
acacactgat gctcaaatcc taagggtcca agctctaggc cctggaggct ggtagaacag      180
gatctatgcc tggaaatcctg gcagggatcc ctgtcaagga cctgtgttta agcctgcttc      240
agggtctcag gctgtctctg ctctgtgtct gccaggctg gctgagcggg tggatgggtg      300
gacagaaggg ctccaccaagg attgtggaca tagggtaggc cctggtagca cgggtttcag      360
gctgttatca ctcccttgtt aggaacatag ccagaagcag atgagccagg gtagagggtc      420
ggccccctct ctcatcttcc ctccagctct aaattgtctc cagcgatggg aagaggccag      480
ggactgtaac ccttgtgtctg tgtattctct gagcctctgc tcaactctcag ggccaagcag      540
ctcccaagcc ggggccccct cctggccaaa atctgaggag cagtctaggt tacaggcttt      600
ttggtaggta ggtctctggc gctgttaaat gcagttaggc cccctgatta ggtacagtga      660
gaaacaagct agaacaaccc tggcccagaa gactgtgcac tccagcaaga tccagggatg      720
atagccttgc agggccactg ggagtttgtg cccaagcttc tccctcttct ctccccaggg      780
ggcactggga ctgggtccctg cctcatcctt tagcctgggc ctccccaga ggtattaaag      840
agaagtatga ttcctctgtc ttcagttctt ttcaggggca tccgtgccat agtaccagt      900
tcccaagggg cccccagtca cgtgggtgaag cctagcactc atgcagctct tagggaaacca      960
aaaaccagca ctgaaataaa gctgaatgac tgactgaaaa aaaaaaaaaa aaagggcggc      1020
cg

```

<210> 76
 <211> 1184
 <212> DNA
 <213> Homo sapiens

<400> 76
 agcaaaaggt gaagagagaa aggaagcttt ttctctaaaa atgggtgcagc tatcctctga 60
 accaattttcc ttcggttttaa tgtacctgta tcttggggtt tttttccact taattttatcc 120
 tggagctctt tccataacaa cacttggaag gcactctcat ccttttttca ctgctgaaca 180
 gaattccact gtgtggatgg aacatactct atttcaccag tccctgtag ccagtcactt 240
 ggtttgtttc caatcttttg ctttttcaga gtaataacct tgtatgtcta tcattttgta 300
 tgcatacagg tttatatgta ggaaaaattc ctagagtagg attgctggac caatggataa 360
 aagtatatgt tggacagaca atgccaattt gcctttcaga gactgtggcc ctgtgcaccc 420
 catcaggcat gtgtgactac caaagctcct gtcagctgtt ttattttatc tcctttccag 480
 tctcaggctc aatgcagaac tttgaggtaa gcttttctaa aatgtaggct cctaaacgcc 540
 acagccagct ctgccacatg aaggagagct caaatgagac agaaacagcc tctgggcagg 600
 atttctatcc ttccacagata tattttccac attctgggaa accgtgaagc ttccagagcc 660
 acaattcccc agaaacacat cccctgtctg gtacagccaa gccccagaac aagctgtgct 720
 tgcttggcac ctcaaagcca agcaccatgg atgccacttg ccatgggtgc ctgcaatttc 780
 aaataatgag aaataagaaa ttccagcttc tcagtccttc tagccaacat ttcagggtgca 840
 tgacagcctc aggtggcaag cagctactct gccggacagg gcagaagatg gaacacccca 900
 tccctgggga agatcacgag gtcaggagat caagaccatc ctggctaaca cagtgaacc 960
 ccgtctctac tgaataatata aaacattagc tgggtgtggt ggcaggcacc tgtagtcca 1020
 gctactcagg aggctgaggc aggagaatgg cgtgaacctg ggaggtggag gtgcagtga 1080
 gccaaagatt cgccactgca ctccagcctg gcaacagagt aagactccat atcaaaaaaa 1140
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 1184

<210> 77
 <211> 312
 <212> DNA
 <213> Homo sapiens

<400> 77
 aattccccgg tgcaccacag cgtccgtgat gagtggattt gtactcttac ccaggctctg 60
 agggccagcc caccagcat cccaccctt gatgacgctg tccctacaac tggctgaact 120
 ggtgcatttt gtgtgtgctt tccagagcca gtggactggt gtgtatccaa tgatgccacc 180
 tctgaaacct acagaaccac tatgctttgc atgtgtacct tgcagggtct gagggccagg 240
 ctgtctggta gctctgctcc tgggtgacag agcaagactc tgtctcaaaa aaaaaaaaaa 300
 agggcgggcg ct 312

<210> 78
 <211> 1370
 <212> DNA
 <213> Homo sapiens

<400> 78
 tgggttaaaga gtacgacatt ttagttaggg ctgagcacag cgcctggccg aggagtgcct 60
 ttatatgctc ttggaagtat atggggactc catctctgtc actgtggcga ttccacttat 120
 gcattcacca tgccaatcca aggcagctga tggcttagga aagtcagaga ctgagatggt 180
 aaaatccttg gggctatcta ccaacatgtc tccattccac ctgctagggt taaaggtttt 240
 tctaaccttg gccctgacct tagcacagat ctgcctatat ttttttgaag ttcagccact 300
 tggactatta gccctaaact tttctgtac tggcactgca gggctgaagg agctttgcat 360
 gcattcacca agtctggctt tcacacctga atttcacacc tcgctttcac ccttagctat 420
 tccattcttc tgtggaacat cagtgtcact tagcaatagc catacaatcc cattatcctt 480
 atacctacct ttcccttcaa agtctcgat gcctgataca ttgcacctgc tagtgcatct 540
 actcccatga gtacactccc aagtccttcc agtgaaagat gtaacaattg aatggccact 600
 gtgccaaggt tgcttgggct ctacctgcca ccagtgatgg ggtcctcaca gccaaaccag 660

tggtgtatca	tccatttcca	aaaccatctc	amtgcctgtt	ttcttggacc	actcctggca	720
tcaactgttt	caggttagag	tgactgaaaa	tttgggktat	aagatattta	ttagagatca	780
atatctatta	aaaatgtgaa	aggaagcagg	attktgctga	ggaagaagat	aacaacaaag	840
ataccacaat	gtcagcccat	caaagccttt	ggccaaccca	gcasaaaaat	ctggagcagg	900
tgtayctttt	tcagagtctc	cccagttagg	tcaaaatgtg	cgcctttaca	cccgcacttc	960
cctcaatcac	gggcttcagg	ctgtcctggg	catgacctca	gatgaagcgg	ctcacacagc	1020
tgaggctaac	gttgtcggag	ctgacagctg	aaagccgttt	gctgaccaca	ctcccacagc	1080
cgagcagcat	gcccttgctt	ggaggaggat	ttggatgaca	cagctctatg	tctgccgtat	1140
tttgggggtc	acattcctca	caccagcttc	tgctcaaat	ggcaacattg	acaaaaatc	1200
attccagttg	caaaccaatt	gtagaatact	ataaaaccag	agtacaagtc	taatagcata	1260
aatctcatcc	taagtggaaa	agaaagggtc	atttatccac	acatttttgg	ggaaaaaaaa	1320
acaacctttg	ctatgtcttt	attacaacac	ggatactcac	aaaaaatagt		1370

<210> 79
 <211> 368
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (5)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (13)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (35)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (351)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (367)
 <223> n equals a,t,g, or c

<400> 79						
aatgnaactc	tctttactaa	ttggaaccaa	agctngagct	ccaccgcggt	ggcggccgct	60
ctagaactag	tggatcccc	gggctgcagg	aattcggcac	gaggtccagg	tgtctttatt	120
tcagatgttc	tgtttttctt	ctattttctg	ctcacatgaa	catacacact	tgccaggcac	180
attctggctt	ttcttatttc	tctttctgat	tttgctctct	tcttgccctt	gttttcttcc	240
cttttctctg	gctatagaaa	ctgtcaggtg	gccttgctgg	catcatccta	cctcttttga	300
actgtgttac	ccaggaactt	ccattttatta	tgctctgagg	gggggcccgg	nacccaattc	360
ggaagant						368

<210> 80
 <211> 1088
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (1)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (4)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (5)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (9)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (11)
 <223> n equals a,t,g, or c

<400> 80
 ntttnnaacng naaatctctcc cttactattg ggaacaaaag ctggagctcc accgcggtgg 60
 cggccgctct agaactagt gatcccccg gctgcaggaa ttcggcacga gattccccag 120
 aattggttca tcttgggaaa ggaaggccta ggaagctgat gacctacttg ttttgttcat 180
 ccatttctct actgcttctt aaagtgcact caagcgggtca ccaggacatc agaaaagcca 240
 aatccaaggt cccaaggctc ttgatcatcc agtgcacaca acagagggag taagctagac 300
 ggagaagctg aagtgaagga ggaaagagag aaggggggtgc ctgttccacc ccttctctag 360
 cagctgcctc tgtgggctcc ataatcccca tgcctctctc cccacccac cctccggccg 420
 mccacctctc ctccaccac cccattcatc agtaggaggg cactgtatca gcacagcctt 480
 tcaaggcctg acagttcctt tctgggtgcc ccccatatgt aaccacttaa cccccagcac 540
 aaagggccag cctgaatggg cttggagccg gtggctcgtg cgraggggga ggcctcaata 600
 ggatttgggg agctgggtgt ataaaccgct ctgctccagc agccttcaat ggaggagctg 660
 gagttggggc atgtgccgag gggaggagct cggcagctgc atcccgscct tgaacaatga 720
 tttattcatt ctcttggcac gcacagagaa gacaggctgg ggaaacatgg cccacaaaat 780
 cgccctcagc ctacagctgat tcccacggca atcattgcag ttcagagtgc cccctccac 840
 ttgtaatgtt ttgcctcaag cttgcccttc ctcttctctg taaggaagta gactacatta 900
 cgggaggcct taagtagect taaattcagg caagcatcgt tctttgagtt ggttgcaccc 960
 aaactcaatc tcactctctc aaagctttct ggccaggttt tcttcacaca cccatagaaa 1020
 cagatgcaga caaaaaaaaa aaaaaaaaaa actcgagggg gggcccgga cccaattcgg 1080
 aagatttg 1088

<210> 81
 <211> 1862
 <212> DNA
 <213> Homo sapiens

<400> 81
 ggcacgagct ctgcaaagct cccctctgctc cagcacgtcc tgtatttttt caacttttga 60
 tgtttgcaca agctattatc ttttgcaaga atgtcctgcc ccttctatc catctgttga 120
 agtcaacctt tcttttagggc ccagaggaaa acctatcacc ttctacacct tcccagatgc 180
 ccaccatttg ggttaaaact tgcctctctc aggtttgcca tggctctttc cctctcttaa 240
 agcactgggc ccagcctatg ccattgtgtg tcaccttggc tcttctctcc tactggctgt 300
 agatcttttg gargggagga actggggaat tttcatcctg gggctatggg atagtgcctta 360

cccctaagta	taaaataagc	aagaatcctt	gaattgaaaa	gctccagtga	cttgatagct	420
tccaacagcc	cctaaagatg	gggtagtcca	ctgctcccta	tataggaagg	aagaggttca	480
cagtctcatg	gatttagcag	ccttcgtagc	ctcaggatct	ttcccaatgt	ccccagttct	540
cctgctgggc	cagctgtggg	ctctgagaat	gctcagttcc	ctgggctagg	ctggcttatg	600
gggccctccc	tccctggacct	gggtccatgt	gaccaagtgg	ccagccagcg	agagagccct	660
gggtgatgag	catctggaag	tcagcttgct	tgtctgcccc	aatcacagcc	ctgggctcta	720
ccctcaagta	ccctactcac	atgggtccaag	agggaaaccac	tcttctgtca	ttccctcccc	780
ttctgctcct	catcctaccc	accagctgt	tctgggcaag	atcctgtggg	catgatggaa	840
tcctcccttt	tctctacccat	ccatcccttt	gatcacctaa	tcctrcarct	gccctccaac	900
caaggtgttc	tttctatacc	aaaaaatctg	atcaggccat	cactcctctg	tttaaaacca	960
ttcaacggct	acccactgcc	ctcaggataa	aaactaaaa	tcttaacagg	atttacaagt	1020
tcttcacctg	gcccttttga	gataagagga	tttataatct	tttcaaagcc	aacaagcatc	1080
ctgagacaat	cttaattttt	gcagctgtcc	aagctattta	agattgatga	ggtctgcamt	1140
aagttgttaa	atgggtgatta	atttatctgg	cctgaatggc	akgtgaattc	caagggggtc	1200
atttttctct	accttccttg	ctgcacattg	ggacacagca	aggccaggac	acccttytta	1260
gggtgactgsg	taaccagggt	cattaaggca	gtcagcgag	agaagaagca	gaaaatgggg	1320
aggagtgtg	taaatgaaat	tgccgataat	accaaaggaa	ccccaagaga	ggttcagttc	1380
cccaacaatt	gcagctttgc	tgcaacatgg	cagggtggga	cagggttagg	tttggatctg	1440
ggctgctttg	ccmtggaagt	gggggttacag	ggcaggaga	ggcttaccat	cactgagcat	1500
ctttaattga	ctttgtcatt	ttastgagaa	cttacaatt	gagctattct	tcattgagtt	1560
atatttgga	gcaggttcaa	tacataatac	agtcctccct	ctagaatttc	tgctttatgt	1620
ccacgtgatg	cagcttgttc	atttctgaa	tccttttaat	gacaaaaaca	tttgggaagga	1680
atgtaaagac	tcaaaaagg	gatggcatgt	aaactatagt	tggaacaata	tggtggccatt	1740
taaaagaagt	gttgtccagg	aggtggaggt	tgagctgagc	tgagattgca	ccactgcact	1800
ctagcctggg	ccatagaatg	agactccttc	tcaaaaaaaa	aaaagaaaaa	aaaaaactcg	1860
ag						1862

<210> 82

<211> 1618

<212> DNA

<213> Homo sapiens

<400> 82

ccacgcgtcc	gcaaggagcc	agaggccatg	cagtggctca	gggtccgtga	gtcgccctggg	60
gaggccacag	gacacagggt	caccatgggg	acagccgccc	tgggtcccgt	ctgggcagcg	120
ctcctgctct	ttctcctgat	gtgtgagatc	cctatggctg	agctcacctt	tgacagagct	180
gtggccagcg	atcgccaacg	gtgctgtgac	tctgaggacc	ccctggatcc	tgcccatgta	240
tcctcagcct	cttccctccg	ccgccccac	gccctgctg	agatcagacc	ctacattaat	300
atcaccatcc	tgaagggtga	caaaggggac	ccaggcccaa	tgggcctgcc	aggggtacatg	360
ggcagggagg	gtccccaaag	ggagcctggc	cctcagggca	gcaagggtga	caagggggag	420
atgggcagcc	ccggcgcccc	gtgccagaag	cgcttctctg	cttcttcagt	ggccgcgaag	480
acggccctgc	acagcggcga	ggacttccag	acgctgctct	tcgaaagggt	ctttgtgaac	540
cttgatgggt	gctttgacat	ggcgaccggc	cagtttggctg	ctcccttgcg	tgcatctac	600
ttcttcagcc	tcaatgtgca	cagctggaat	tacaaggaga	cgtacgtgca	cattatgcat	660
aaccagaaag	aggctgtcat	cctgtacgag	cagcccagcg	agcgcagcat	catgcagagc	720
cagagtgtga	tgctggacct	ggcctacggg	gaccgcgtct	gggtgcggct	cttcaagcgc	780
cagcgcgaga	acgccatcta	cagcaacgac	ttcgacacct	acatcacctt	cagcggccac	840
ctcatcaagg	ccgaggacga	ctgagggcct	ctgggccacc	ctcccggtcg	gagagctcag	900
ctgatacggc	atcctgcgag	aagacctgcc	ctcctcactg	ggatccccct	cctgcctcct	960
cccagggctc	tgccagggcc	ttgctcagtc	cttccaacca	aagtcattctg	aacttccgtt	1020
tcccagggcc	tccagctgcc	ctcagacact	gatgtctgtc	cccaggtgct	ctctgccccct	1080
catgccccct	tcaccggccc	agtgcctcga	ctctccaggc	tttatcaagg	tgctaaggcc	1140
cgggtgggca	gtcctctcgt	tcagagccct	cctccggcct	ggtgctgcct	ttacaaacac	1200
ctgcaggaga	agggccacgg	aagccccagg	ctttagagcc	ctcagcaggt	ctggggagct	1260
agagcaaagg	agggacctca	ggccttccgt	ttcttcttcc	aggggggggt	ggcctgggtgt	1320
tcctctagct	ttccaaacct	aggtggcctg	cccttctccc	cagagggagg	cggcctccgc	1380
ccattgggtc	tctgcagac	tctggggctg	aggtgccccg	gggggtgac	tctgggtgctc	1440
acagtcgagg	gagccgtggc	tccatggcca	gatgacggaa	acagggctctg	accaagtgcc	1500

aggaagacct gtgctataaa ccaccctgcc tgatectgcc cctgcctgac cccgccacgc 1560
cctgccgtcc agcatgatta aagaatgctg tctcctcttg gaaaaaaaaa aaaaaaaa 1618

<210> 83
<211> 2034
<212> DNA
<213> Homo sapiens

<220>
<221> SITE
<222> (14)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (382)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (1999)
<223> n equals a,t,g, or c

<220>
<221> SITE
<222> (2027)
<223> n equals a,t,g, or c

<400> 83
actcactaag ggancaaagc tggagctcca ccgcggtggc ggccgctcta gaactagtgg 60
atcccccggg ctgcaggaat tcggcacgag atctaccggg agakggaytc tgaaagggcc 120
ccggccagcg tccctgagac gccaacggca gtcactgccc cccattccag ctccctgggat 180
acgtactatc agccccgtgc cctggagaaa catgctgaca gcatcctggc actggcttca 240
gtattctggt ccatctctta ttactcctct cccttcgect tcttctactt gtacaggaaa 300
ggttacttga gtttgtccaa agtgggtgccg ttttctcact atgctgggac attgctgcta 360
cttctggcag gtgtggcctg cntccgaggc attggcctct ggaccaacce ccagtaccgg 420
cagttcatca ccatcttggg agcaacacat cggaaccagt cttcagaaaa caagaggcag 480
cttgccaact acaactttga ctcccgagc tggccagtcg acttccactg ggaagaaccc 540
agcagccgga aggagtctcg agggggccct tcccgccggg gtgtggccct gcttcgcca 600
gagccctgc accgggggac agcagacacc ctcccaacc gggtaagaa gctgccttgt 660
cagatcacca gctacctggt ggccgacacc ctaggcgccc ggatgctgta tccaggctct 720
gtgtacctgc tgcagaaggc cctcatgcct gtgctgctgc agggccaggc ccgactgggtg 780
gaagagtgtg atgggcgccc ggcaaagctg ctggcctgtg atggcaatga gattgacacc 840
atgtttgtgg accggcgggg gacagctgag ccccgaggac agaagctggt gatctgctgt 900
gaggggaatg ctgggtttta tgaggtgggc tgcgtctcca cggccctgga agctggatat 960
tcagtcctgg gctggaatca tccaggctct gctggaagca cgggggtgcc attcccgag 1020
aatgaggcta atgccatgga tgtggtggtc cagtttgcca tccaccgctc rggcttccag 1080
ccccaggaca tcatcatcta cgccctggtc atcgcgggct tcaactgccac gtgggcagcc 1140
atgtcttacc cagatgttag tgccatgatc ctggatgcct cctttgatga cctggtgccc 1200
ttggccttga aggtcatgcc agacagctgg aggggcctgg tgaccaggac cgtgaggcag 1260
catctcaatc taaacaacgc ggagcagctg tgcagatacc agggctcctgt actgctgatc 1320
cggagaacca aggatgagat catcaccacc acgggttctg aggacatcat gtccaaccga 1380
ggcaatgacc tctgtctgaa gctcctgcag catcggtatc cccgggtgat ggcagaggag 1440
ggtcttctgag tgggtgaggca gtggttggag gcctcctcac agctggaggga agcctcaatt 1500
tatagccgat gggaggttga agaggactgg tgtctgtctg tctccgctc ctaccaggca 1560
gaacacgggc ccgatttccc ctggagcgtg ggggaggaca tgagtgcaga tggacggcgg 1620
cagctggctt tggttctggc tcggaagcat ctgcacaact ttgaggccac tcaactgcacc 1680
cactcccgag cccagaactt ccagatgccc tggcacctct agggaccaac tgggactcat 1740

tatggaagaa	tgggggtgaga	ggagacatga	ggaaagaccc	tcttatttgt	gattctctgt	1800
gttcatgttg	ctgtttatag	tttgtggaaa	gtgggggacc	atcccccttc	tcaccactgt	1860
tcctcttgca	cgtttccctt	cattcatgtg	gctgtactta	accttctcca	acatacatcc	1920
tgcatcatat	gaatggatta	ttcctaataa	ttaataaaaa	ggtatttttt	ctacaaaaaa	1980
aaaaaaaaaa	aaactccana	attacttctt	aaaaccggcc	cgcggncccc	atcc	2034

<210> 84

<211> 2240

<212> DNA

<213> Homo sapiens

<400> 84

tgcaccacgc	gtccgcttcc	tcattgetgcc	cgctcgcccta	ccgttcaggc	tgctgagcct	60
tttccttcgt	ggatccgctc	ccacggcagc	gcgccatggc	ctccgggagc	cgctcctgga	120
gaggaggtgc	gctgccgcct	cctccttcca	gcactcatcg	agtctgggac	gcgagcttcc	180
ttatgacccc	gtggacacgg	agggttttgg	agaaggtggt	gacatgcagg	agcgttttct	240
gttccccgag	tacatcctgg	atccggagcc	gcaaccacc	cgcgaaaagc	agctgcagga	300
gctccagcaa	cagcaggagg	aggaggagcg	acagaggcag	cagcgggcgg	aggagcggcg	360
acagcaaac	ctacgggcca	ggccccggga	gcacccggtc	gtggggcacc	cggacccggc	420
attgccgccc	agcggcgctga	actgctcggg	ctgcggggca	gagctgcact	gccaggacgc	480
cggagtgcgc	ggctacctgc	cccagagaaa	gttccctccg	acggcgagg	cagacggcgg	540
gctggcacgg	accgtgtgcc	agcgtgtctg	gctgctgtcg	caccaccggc	gcgctctacg	600
cctgcaggtg	agcgcgcagc	agtacctgga	gctgggtgagc	gccgcgttgc	ggcggcccg	660
cccctccctg	gtgctctaca	tggtggacct	gctggacctg	cccagcggcc	tgctgcccga	720
cttgcccgcg	ctgggtggcc	ccaagcagct	gatcgtgtcg	ggaaacaaag	tggaacctct	780
gccccaggat	gtccttggtc	accggcagag	gctgcgggag	cgactgtggg	aggactgtgc	840
ccgcgccggg	ctcctgctgg	cccctggcca	ccaagggcca	cagcgccccg	tcaaggacga	900
gccacaggac	ggggagaatc	cgaatccgcc	gaactgggtc	cgcacagtgg	tcagggacgt	960
gcggctgatc	agcgccaaga	ccggctatgg	agtggaaagag	ttgatctctg	cccttcagcg	1020
ctcctggcgc	taccgtgggg	acgtctactt	agtgggcgcc	accaacggcc	gcaaattccac	1080
tctctttaac	acgtcctctg	agtccgatta	ctgcactgcc	aagggtctcg	acgccatcga	1140
cagagccacc	atctccccct	ggccagggtac	tacattaaac	cttctgaagt	ttcctatttg	1200
caaccacaact	ccttacagaa	tggttaaaaag	gcattcaaaa	cttaaaaaag	attcaactca	1260
agctgaagaa	gatcttagtg	agcaagaaca	aaatcagctt	aatgtcctca	aaaagcatgg	1320
ttatgtcgta	ggaagagttg	gaaggacatt	cttgatttca	gaagaacaga	aggataacat	1380
tcctcttgag	tttgatgctg	attcacttgc	ctttgacatg	gaaaatgacc	ctgttatggg	1440
tacacacaaa	tccaccaaac	aagtagaatt	gactgcacaa	gatgtgaaag	atgccactg	1500
gttttatgac	acccttgga	ttacaaaaga	aaattgtatt	ttaaatcttc	taacagaaaa	1560
agaagttaa	attgttttgc	caacacagtc	cattgttcca	agaacttttg	tgcttaaacc	1620
aggaatggtt	ctgttttttg	gtgctatagg	ccgcatagat	ttcctgcagg	gaaatcagtc	1680
agcttggttt	acagtctgtg	cttccaacat	cctccctgtg	catatcacct	ccttggaacg	1740
ggcagacgct	ctgtatcaga	agcatgcagg	tcatacgtta	ctccagattc	caatgggtgg	1800
aaaagaacga	atggcaggat	ttcctctctt	tgttgctgaa	gacattatgt	taaaagaagg	1860
actgggggca	tctgaagcag	tgggcgacat	caagttttcc	tctgcagggt	gggtttcagt	1920
aacaccta	tttaaggaca	gactgcattc	ccgaggctat	acacctgaag	gaacagtttt	1980
gaccgtccgg	ccccctctct	tgccatatat	tgtaaacatc	aaaggacagc	gcattcaagaa	2040
aagtgtggcc	tataaaacca	agaagcctcc	ttcccttatg	tacaacgtga	ggaagaagaa	2100
aggaaagata	aatgtatgag	accgaccttg	ttcactccag	atatttaactg	tattgaacac	2160
aacaaaatac	attgaatttg	tattaaacat	ataacgcata	aataaagctc	ccattctttac	2220
ccttaaaaaa	aaaaaaaaaa					2240

<210> 85

<211> 1488

<212> DNA

<213> Homo sapiens

<400> 85

cgccaagttt	cgggagggag	agggtagaaa	ctggaggggg	tggacctgtc	actcacggga	60
ctgaggggtcc	ttttctcccg	ctcccaggag	gaacgagaat	gaatatgact	caagcccggg	120
ttctgggtggc	tgcagtgggtg	gggttgggtg	ctgtcctgct	ctacgcctcc	atccacaaga	180
ttgaggagggg	ccatctggct	gtgtactaca	ggggaggagc	ttactaact	agccccagtg	240
gaccaggcta	tcatatcatg	ttgcctttca	ttactacgtt	cagatctgtg	cagacaacac	300
tacaaactga	tgaagttaaa	aatgtgcctt	gtggaacaag	tgggtggggtc	atgatctata	360
ttgaccgaat	agaagtgggtt	aatatgttgg	ctccttatgc	agtgtttgat	atcgtgagga	420
actatactgc	agattatgac	aagaccttaa	tcttcaataa	aatccaccat	gagctgaacc	480
agttctgcag	tgcccacaca	cttcaggaag	tttacattga	attgtttgat	caaatagatg	540
aaaacctgaa	gcaagctctg	cagaaagact	taaacctcat	ggccccagggt	ctcactatac	600
aggctgtgcg	tgttacaaaa	cccaaaatcc	cagaagccat	aagaagaaat	tttgagttaa	660
tggaggctga	gaagacaaaa	ctccttatag	ctgcacagaa	acaaaagggt	gtggaaaaag	720
aagctgagac	agagaggaaa	aaggcagtta	tagaagcaga	gaagattgca	caagtggcaa	780
aaattcgggtt	tcagcagaaa	gtgatggaaa	aagaaactga	aaagcgcatt	tctgaaatcg	840
aagatgctgc	attcctggcc	cgagagaaaag	cgaaagcaga	tgctgaatat	tatgctgcac	900
acaaatatgc	cacctcaaac	aagcacaaagt	tgacccccgga	atatctggag	ctcaaaaagt	960
accaggccat	tgttcttaac	agtaagatct	atattggcag	caacatccct	aacatgttcg	1020
tggactcctc	atgtgctttg	aaatattcag	atattaggac	tggaaagaaa	agctcactcc	1080
cctctaagga	ggctcttgaa	ccctctggag	agaacgtcat	ccaaaacaaa	gagagcacag	1140
gttgatgcaa	gaggtggaaa	tgttctccat	atcaagatgt	ggcccaagggt	gttaagtggg	1200
aacaatcatt	atacggactc	ttcagattta	cagagaactt	acacttcac	tggtccacct	1260
ctcctgcgat	agtcctgggt	gctccactga	ttggaggata	gagccagctg	tctgacacac	1320
aaatgggtctt	ttcagccaca	gtcttatcaa	gtatcctata	tgtattcctt	tctaaactgc	1380
tactcatgaa	tgaggaaaagt	ctgatgctaa	gatactgcct	gcactggaat	gttaaacact	1440
aaatatataa	caagctgtgt	tttcctaagc	tgaaaaaaaaa	aaaaaaaaa		1488

<210> 86

<211> 3174

<212> DNA

<213> Homo sapiens

<400> 86

gcggacgctg	grcscaaaca	ctaaggcctg	agcgggtgaca	atcgaggcga	gatgatgggtc	60
aacagggaat	gcctcgtggg	agaaaaaaga	caattttatt	ctcagcgctg	atttttgagat	120
gatgggcttg	ggaaaacgggc	gtcgcagcat	gaagtcgccg	cccctcgtgc	tggccgcctt	180
ggtggcctgc	atcatcgtct	tgggcttcaa	ctactggatt	gcgagctccc	ggagcgtgga	240
cctccagaca	cggatcatgg	agctggaagg	cagggtccgc	aggcgggctg	cagagagagg	300
cgccgtggag	ctgaagaaga	acgagttcca	gggagagctg	gagaagcagc	gggagcagct	360
tgacaaaatc	cagtcagacc	acaacttcca	gctggagagc	gtcaacaagc	tgtaccagga	420
cgaaaaggcg	gttttggtga	ataacatcac	cacaggtgag	aggctcatcc	gagtgtgca	480
agaccagtta	aagaccctgc	agaggaatta	cggcaggctg	cagcaggatg	tcctccagtt	540
tcagaagaac	cagaccaacc	tggagaggaa	gttctcttac	gacctgagcc	agtgcatcaa	600
tcagatgaag	gaggtgaagg	aacagtgtga	ggagcgaata	gaagaggcca	ccaaaaagggt	660
gaatgaagct	gtagcttcca	gagacctgag	tgaaaacaac	gaccagagac	agcagctcca	720
agccctcagt	gagcctcagc	ccaggctgca	ggcagcaggc	ctgccacaca	cagaggtgcc	780
acaagggaag	ggaaacgtgc	ttggtaacag	caagtcccag	acaccagccc	ccagttccga	840
agtggttttg	gattcaaaga	gacaagtga	gaaagaggaa	accaatgaga	tccaggtgggt	900
gaatgaggag	cctcagaggg	acaggctgcc	gcaggagcca	ggccgggagc	aggtgtgtgga	960
agacagacct	gtaggtggaa	gaggcttcgg	gggagccgga	gaactgggcc	agacccaca	1020
ggtgcaggct	gccctgtyag	tgagccagga	aaatccagag	atggagggcc	ctgagcgaga	1080
ccagcttgct	atccccgacg	gacaggagga	ggagcaggaa	gctgccgggg	aaggagaaaa	1140
ccagcagaaa	ctgagaggag	aagatgacta	caacatggat	gaaaaatgaag	cagaatctga	1200
gacagacaag	caagcagccc	tggcagggaa	tgacagaaac	atagatgttt	ttaatgttga	1260
agatcagaaa	agagacacca	taaatttact	tgatcagcgt	gaaaagcggg	atcatacact	1320
ctgaattgaa	ctggaatcac	atatttcaca	acagggccga	agagatgact	ataaaatgtt	1380
catgagggac	tgaatactga	aaactgtgaa	atgtactaaa	taaaatgtac	atctgaagat	1440
gattattgtg	aaattttagt	atgcactttg	tgtaggaaaa	aatggaatgg	tcttttaaac	1500
agctttttggg	gggtactttg	gaagtgtcta	ataaggtgtc	acaattttctg	gtagtaggta	1560

tctcgtgaga	agttcaacac	caaaactgga	acatagttct	ccttcaagtg	ttggcgacag	1620
cggggcttcc	tgattctgga	atataacttt	gtgtaaatta	acagccacct	atagaagagt	1680
ccatctgctg	tgaaggagag	acagagaact	ctgggttccg	tcgtccctgtc	cacgtgctgt	1740
accaagtgt	ggtgccagcc	tgttacctgt	tctcactgaa	aagtctggct	aatgctcttg	1800
tgtagtcact	tctgattctg	acaatcaatc	aatcaatggc	ctagagcact	gactgttaac	1860
acaaacgtca	ctagcaaagt	agcaacagct	ttaagtctaa	atacaaagct	gttctgtgtg	1920
agaatttttt	aaaaggctac	ttgtataata	acccttgtca	tttttaattg	acaaaacgct	1980
attaagtggc	ttagaatttg	aacatttgtg	gtctttattt	actttgcttc	gtgtgtgggc	2040
aaagcaacat	cttcctaaa	tatatattac	caagaaaagc	aagaagcaga	ttaggttttt	2100
gacaaaacaa	acaggccaaa	agggggctga	cctggagcag	agcatggtga	gaggcaaggc	2160
atgagagggc	aagtttggtg	tggacagatc	tgtgcctact	ttattactgg	agtaaaagaa	2220
aacaaagtgc	attgatgtcg	aaggatatat	acagtgttag	aaattaggac	tgtttagaaa	2280
aacaggaata	caatgggtgt	ttttatcata	gtgtacacat	ttagcttgtg	gtaaatgact	2340
cacaaaactg	attttaaaat	caagttaatg	tgaattttga	aaattactac	ttaatcctaa	2400
ttcacaataa	caatggcatt	aaggtttgac	ttgagttggt	tcttagtatt	atttatggta	2460
aataggtctt	taccacttgc	aaataactgg	ccacatcatt	aatgactgac	ttcccagtaa	2520
ggctctctaa	ggggtaagta	ggaggatcca	caggatttga	gatgctaagg	ccccagagat	2580
cgtttgatcc	aaccctctta	ttttcagagg	ggaaaatggg	gcctagaagt	tacagagcat	2640
ctagctgggt	cgctggcacc	cctggcctca	cacagactcc	cgagttagctg	ggactacagg	2700
cacaggtca	ctgaagcagg	ccctgtttgc	aattcacgtt	gccacctcca	acttaaacat	2760
tcttcatatg	tgatgtcctt	agtcactaag	gttaaaacttt	cccaccacaga	aaaggcaact	2820
tagataaaat	cttagagtac	tttcatactc	ttctaagtcc	tcttccagcc	tcactttgag	2880
tcctccttgg	ggttgatagg	aattttctct	tgctttctca	ataaagtctc	tattcatctc	2940
atgtttaatt	tgtacgcata	gaattgctga	gaaataaaat	gttctgttca	acttaaaaaa	3000
cttgtcacag	ctctgttttt	ttgcttttcc	ccttcactct	ctattagatc	cttaagaagt	3060
ccaccctcaa	actgtggact	aaggatggca	gtttgcctta	aaccttctgt	ggaatgagat	3120
gggctgtcag	gacgtcttca	accacagta	ccccgaaagc	tttgtgtcga	ccac	3174

<210> 87

<211> 2780

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (2760)

<223> n equals a,t,g, or c

<400> 87

ggcgccgcgg	gcgggggggtc	ccggggccac	gtggacgccg	acggctctgg	tctgcggcgt	60
gtggacactg	ccagcggctg	cagccgactt	ggaatgacct	gggagacaaa	tacaacagca	120
tggaaagakgc	caaagtctat	gtggctaaag	tggactgcac	ggccccactcc	gacgtgtgct	180
ccgcccaggg	ggtgcgagga	taccccacct	taaagctttt	caagccaggc	caagaagctg	240
tgaagtacca	gggtcctcgg	gacttccaga	cactggaaaa	ctggatgctg	cagacactga	300
acgaggagcc	agtgcacca	gagccggaag	tggaaaccgcc	cagtgcctccc	gagctcaagc	360
aagggtctga	tgagctctca	gcaagcaact	ttgagctgca	cgttgcacaa	ggcgaccact	420
ttatcaagt	cttcgctccg	tggtgtgggc	actgcaaagc	cctggctcca	acctgggagc	480
agctggctct	gggccttgaa	cattccgaaa	ctgtcaagat	tggcaagggt	gattgtacac	540
agcactatga	actctgctcc	ggaaaccagg	ttcgtggcta	tcccactctt	ctctgggttc	600
gagatgggaa	aaaggtggat	cagtacaagg	gaaagcggga	tttggagtca	ctgaggaggt	660
acgtggagtc	gcagctgcag	cgcacagaga	ctggagccac	ggagaccgtc	acgccctcag	720
aggccccggg	gctggcagct	gagcccgagg	ctgacaaggg	cactgtgttg	gcactcactg	780
aaaataactt	cgatgacacc	attgcagaag	gaataacctt	catcaagttt	tatgctccat	840
ggtgtgggtca	ttgtaaagact	ctggctccta	cttgggagga	actctctaaa	aaggaattcc	900
ctggtctggc	gggggtcaag	atcgccgaag	tagactgcac	tgctgaacgg	aatatctgca	960
gcaagtattc	ggtacgaggc	taccccacgt	tattgctttt	ccgaggaggg	aagaaagtca	1020
gtgagcacag	tggaggcaga	gaccttgact	cgttacaccg	ctttgtcctg	agccaagcga	1080
aagacgaact	ttaggaacac	agttggagggt	cacctctcct	gcccagctcc	cgcaccctgc	1140

gtttaggagt	tcagtccac	agaggccact	gggttcccag	tgggtggctgt	tcagaaagca	1200
gaacatacta	agcgtgaggt	atcttctttg	tgtgtgtgtt	ttccaagcca	acacactcta	1260
cagattcttt	attaagttaa	gtttctctaa	gtaaagtgtt	aactcatggt	cactgtgtaa	1320
acattttcag	tggcgatata	tcccccttga	ccttctcttg	atgaaattta	catggtttcc	1380
tttgagacta	aaatagcgtt	gagggaaaatg	aaattgctgg	actatttgtg	gctcctgagt	1440
tgaagtgaatt	tggtgaaaga	aagcacatcc	aaagcatagt	ttacctgccc	acgagttctg	1500
gaaaggtggc	cttgtggcag	tattgacgtt	cctctgatct	taaggtcaca	gttgactcaa	1560
tactgtgttg	gtccgtagca	tggagcagat	tgaaatgcaa	aaaccacac	ctctggaaga	1620
taccttcacg	gccgtgctg	gagcttctgt	tgctgtgaat	acttctctca	gtgtgagagg	1680
ttagccgtga	tgaaagcagc	gttacttctg	accgtgcctg	agtaagagaa	tgctgatgcc	1740
ataactttat	gtgtcgatac	ttgtcaaate	agttactgtt	caggggatcc	ttctgtttct	1800
cacgggggtga	aacatgtctt	tagttcctca	tgttaacacg	aagccagagc	ccacatgaac	1860
tgttggatgt	cttccctaga	aagggtaggc	atggaaaatt	ccacgaggct	cattctcagt	1920
atctcattaa	ctcattgaaa	gattccagtt	gtatttgtca	cctgggggtga	caagaccaga	1980
caggctttcc	caggcctggg	tatccaggga	ggctctgcag	ccctgctgaa	gggccctaac	2040
tagagttcta	gagtttctga	ttctgtttct	cagtagtcct	tttagaggct	tgctataact	2100
ggtctgcttc	aaggaggctg	accttctaata	gtatgaagaa	tgggatgcat	ttgatctcaa	2160
gaccaaagac	agatgtcagt	gggctgctct	ggcctctggg	tgcacggctg	tggcagctgt	2220
tgaatgccagt	gtcctctaac	tcattgctgtc	cttgtgatta	aacacctcta	tctcccttgg	2280
gaataagcac	atcacaggctt	aagctctaag	ataggtgttt	gtcctttttac	catcgagcta	2340
cttcccataa	taaccacttt	gcaccaaca	ctcttcaccc	acctcccata	cgcaagggga	2400
tgtggatact	tggcccaaag	taactggtgg	taggaatctt	agaaacaaga	ccacttatac	2460
tgtctgtctg	aggcagaaga	taacagcagc	atctcgacca	gcctctgcct	taaaggaaat	2520
ctttattaat	cacgtatggt	tcacagataa	ttcttttttt	aaaaaaaacc	aacctcttag	2580
agaagcacaa	ctgtcaagag	tcttgtacac	acaacttcag	ctttgcatca	cgagctctgt	2640
attccaagaa	aatcaaaagt	gtacaatttg	tttgtttaca	ctatgatact	ttctaataaa	2700
actctttttt	tttaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaggggg	gggggggggn	2760
ggggcccccc	cccccccaaa					2780

<210> 88

<211> 1061

<212> DNA

<213> Homo sapiens.

<400> 88

aattcggcac	gagagaagga	aatacatcaa	aatgcccaca	ttggttatct	gtagaggata	60
gaatgaaaga	tggcttttatt	ctcttgtttg	ctcttattaa	agcaatcaga	tgggtgcttct	120
cctgtactca	gagccctggc	tgttctctgc	ctggcctctc	ctgggggctg	ctgtggaacc	180
agaaaagcct	taaacggaaa	tgtgggagag	aagggtggat	tcactttcat	gtctttccag	240
ggttgtgacc	cctcaagtcc	tggttgcctt	tgtgtgtctc	tattaccttc	aaacagccag	300
ctcgtcttta	tttctttttt	agttttgtcg	gggttggctt	gatagatgtt	agtcacatcat	360
agccagatgt	gtctagcctt	gtcttttgaa	tgcaagattt	aggatgtggg	tacttagctg	420
ttagtggaca	tcagagtcac	tagtcaggat	gaaagagttc	ttggctttaa	ctcccagaaa	480
ttctggtaac	gtcatgtata	gtgacggccg	catgtctaac	agggtggccag	gtaagtcttt	540
tgggggtggc	tgtgaatcac	agtttgggag	acattgactt	ttaggagatt	tgttctgaat	600
tcactagata	atagagatat	aatacagagc	tttgaaagct	gggtgtcttg	tgacagagcc	660
gtggcaatgg	ggagggttga	ggaggtggct	gttgggcctg	tctcctggtg	agagttgaaa	720
gggcctgaac	tcaagcagag	gcctcagaac	cgaaagggtg	tgggaaggatg	cagcaagagg	780
cgccacacag	gagtactctg	cgccctggga	gggtctgaat	acacgtggga	gtgggtgagag	840
ggagaacttt	aagtcagggt	tttgtgcctc	agtgaactag	tgtggccata	tcattagaaa	900
tgtgttgagg	ccgggcacag	tggctcatgt	ctgtaattcc	agcactttga	gaggctgagg	960
caggaggatg	gcttgaggcc	aggagttaa	aaccagcctg	gacaacatag	tgagagcctg	1020
tctctacaaa	aaaaaaaaaa	aaaaactcga	ggggggggccc	g		1061

<210> 89

<211> 1342

<212> DNA

<213> Homo sapiens

<400> 89

ccacgcgtcc	ggcgcgcgcg	gtgaaaggcg	cattgatgca	gcctgcggcg	gcctcggagc	60
gcggcgagc	agacgctgac	cacgttcctc	tcctcggtct	cctccgcctc	cagctccgcg	120
ctgcccggca	gccgggagcc	atgcgacccc	aggggcccg	cgcctccccg	cagcggctcc	180
gcggcctcct	gctgctcctg	ctgctgcagc	tgcccgcgcc	gtcgagcgcc	tctgagatcc	240
ccaaggggaa	gcaaaaggcg	cagctccggc	agagggaggt	ggtggacctg	tataatggaa	300
tgtgcttaca	agggccagca	ggagtgcctg	gtcgagacgg	gagccctggg	gccaatggca	360
ttccgggtac	acctgggatc	ccaggctcgg	atggattcaa	aggagaaaag	ggggaatgtc	420
tgagggaag	ctttgaggag	tcctggacac	ccaactacaa	gcagtgttca	tggagtccat	480
tgaattatgg	catagatcct	gggaaaattg	cggagtgtac	atttacaag	atgcgttcaa	540
atagtgtct	aagagttttg	ttcagtggct	cacttcggct	aaaatgcaga	aatgcatgct	600
gtgagcggtg	gtatttcaca	ttcaatggag	ctgaatgttc	aggacctctt	cccattgaag	660
ctataattta	tttgaccac	ggaagccctg	aaatgaattc	aacaattaat	attcatcgca	720
cttcttctgt	ggaaggactt	tgtgaaggaa	ttggtgctgg	attagtggat	gttgctatct	780
gggttggcac	ttgttcagat	taccacaaag	gagatgcttc	tactggatgg	aattcagttt	840
ctcgcatcat	tattgaagaa	ctaccacaa	aaatgcttta	attttcattt	gctacctctt	900
tttttattat	gccttggaa	ggttcactta	aatgacattt	taataagtt	tatgtataca	960
tctgaatgaa	aagcaaagct	aaatatgttt	acagacacaa	gtgtgatttc	acactgtttt	1020
taaatctagc	attattcatt	ttgcttcaat	caaaagtggt	ttcaatat	tttttagttg	1080
gttagaatac	tttcttcata	gtcacattct	ctcaacctat	aatttggaa	attgttgtgg	1140
tcttttgttt	tttctcttag	tatagcattt	ttaaaaaat	ataaaagcta	ccaatctttg	1200
tacaatttgt	aaatgttaag	aatttttttt	atatctgtta	aataaaaatt	atttccaaaa	1260
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aa				1342

<210> 90

<211> 770

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (690)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (762)

<223> n equals a,t,g, or c

<400> 90

ggtagggtcg	ggkcgaccyr	cgcgccggc	tcttcctcag	gcggcgggcca	tggcgggaca	60
ggaggatccg	gtgcagcggg	agattcacca	ggactgggct	aaccgggag	acattgagat	120
aatcaccagc	agcatcaaga	aaatcgcaga	ctttctcaac	tcgttcgata	tgtcttgctg	180
ttcaagactt	gcaacactaa	acgagaaatt	gacagccctt	gaacggagaa	tagagtacat	240
tgaagctcgg	gtgacaaaag	gtgagacact	cacctagaac	agtgcctg	tgctgctggg	300
aagttgtctt	acacaacaca	ggccacatgg	gaaaggcccc	agcagccttc	agctccttcc	360
tttctcctta	aagagcaaca	gggcttatcc	ttgtttttct	tttttcaaaa	gtgtggcctt	420
tgggctctgc	catctggggg	gtggtgtggg	atgtgggaag	aagttcagag	gaaccgttgg	480
aaacgacgtt	aggcatttta	ccttttcagt	aacattttat	acatctactt	gtcaatgtat	540
ttgagacatt	cacagccaaa	agcctgggac	tctttgtgaa	ggtcctcctc	amctctatct	600
ttctttctct	ctctctcaaa	ctttccttaa	agttctcatt	gcctttgcac	tgcttctgtg	660
aacagtcttt	gtctcctccc	cacctttggg	ggaaagtgcg	gggcagtcct	ggtcaagaca	720
ctcatgcctt	ggcaatgtgg	cctgcagaga	atgttgttgt	anccaccagt		770

<210> 91
 <211> 1570
 <212> DNA
 <213> Homo sapiens

<400> 91
 gatgggtttta ctgaagttgc taaaagttta cagaaaagga agtgcaggaa catttcacaa 60
 atctacaatc tgtgagtatc acatccctgt tagctgtmaa cactggaata aggaagggct 120
 gatgactttc agaagatgaa ggtaagtaga aaccgttgat gggactgaga aaccagagtt 180
 aaaacctctt tggagcttct gaggactcag ctggaaccaa cgggcacagt tggcaacacc 240
 atcatgacat cacaaccctgt tcccaatgag accatcatag tgctcccatc aaatgtcatc 300
 aactttctccc aagcagagaa acccgaaccc accaaccagg ggcaggatag cctgaagaaa 360
 catctacacg cagaaatcaa agttattggg actatccaga tcttgtgtgg catgatggta 420
 ttgagcttgg ggatcatttt ggcatctgct tcttctctc caaattttac ccaagtgact 480
 tctacactgt tgaactctgc ttacccattc ataggaccct ttttttttat catctctggc 540
 tctctatcaa tcgccacaga gaaaaggtr accaagcttt tgggtgcatag cagcctgggt 600
 ggagcattc tgagtgtctt gtctgccttg gtgggtttca ttatcctgtc tgtcaaacag 660
 gccaccttaa atcctgcctc actgcagtgt gaggctggaca aaaataatat accaacaaga 720
 agttatgttt cttactttta tcatgattca ctttatacca cggactgcta tacagccaaa 780
 gccagtctgg ctggawctct ctctctgatg ctgatttgca ctctgtctga attctgccta 840
 gctgtgtctca ctgctgtgct gcggtggaaa caggcttact ctgacttccc tgggagtgta 900
 cttttctctg ctacagtta cattggtaat tctggcatgt cctcaaaaat gactcatgac 960
 tgtggatag aagaactatt gacttcttaa gaaaaaagg agaaatatta atcagaaagt 1020
 tgattcttat gataatatgg aaaagttaac cattatagaa aagcaaagct tgagtttctt 1080
 aaatgtaagc ttttaaagta atgaacatta aaaaaaacca ttatttctact gtcatttaag 1140
 atatgtgttc attggggatc tcttgatttg cctgacattg acttcagcaa aagcacgggg 1200
 ctgtaaatca ccatttacta gattagccaa atagtctgaa tttccagaaa acaaggcaga 1260
 atgatcattc ccagaaacat tcccagaaa atgtttccca gaaaactaga cagmatgatc 1320
 attcaatgga tcacagtga gcaaaggaca caacttttta ttgtaccct taattgtcaa 1380
 caggagttaa ctgatttggt gtggtgctca gactttttta tacaggtgct agtgttttat 1440
 cctatgtatt ttaactcatt agtgcataaa ggcaagcccc atataatgaa gtctcagggt 1500
 atatgaaagt agctggcttc aaaataaaat ttttgagtgc aaaaaaaaaa aaaaataaaa 1560
 aaaaaaaaaa 1570

<210> 92
 <211> 2950
 <212> DNA
 <213> Homo sapiens

<400> 92
 ccgggtctcc gcccgctccc agccggggccc cccagcggtc ggccgggacgg ctcccggtcg 60
 cagtctgccc gcccgccccg cgcggggccc gagtgcgaa gcgcgcctgc gaccggcggt 120
 ccgggcgcgc tggagaggac gcgaggagcc atgaggcgcc agctgcgaag gtggcggcgc 180
 tgctgtctcg gctgtctctg gagtgcacag aagccaaaaa gcattgtctg tatttcgaag 240
 gactctatcc aacctattat atatgccgt cctacgagga ctgctgtggc tccagggtgct 300
 gtgtgcgggc cctctccata cagaggctgt ggtacttctg gttccttctg atgatgggcg 360
 tgcttttctg ctgcggagcc ggcttcttca tccggaggcg catgtacccc ccgcgcctga 420
 tcgaggagcc agccttcaat gtgtcctaca ccaggcagcc cccaaatccc ggcccaggag 480
 ccagtcagcc ggggccgccc tattacacyg acccaggagg accggggatg aacctgtcg 540
 ggaattccat ggcaatggct ttccaggctc cacccaactc accccagggg agtgtggcct 600
 gcccgcccc tccagcctac tgcaacacgc ctccgcccc gtacgaacag gtagtgaagg 660
 ccaagtagtg gggtgccac gtgcaagagg agagacagga gagggcctt cctggcctt 720
 tctgtctctg ttgatgttca cttccaggaa cggctctctg ggtgtctaag ggcagttcct 780
 ctgatatact cacagcaagc acagctctct ttcaggcttt ccatggagta caatatatga 840
 actcacactt tgtctctctt gttgtctctg tttctgacgc atctgtgctc tcacatggta 900
 gtgtgtgac agtccccgag ggctgacgtc cttacggtgg cgtgaccaga tctacaggag 960
 agagactgag aggaagaagg cagtgtgga ggtgcagggt gcatgtagag gggccaggcc 1020
 gagcatccca ggcaagcatc cttctgcccc ggtattaata ggaagcccca tgccgggcgg 1080

ctcagccgat	gaagcagcag	ccgactgagc	tgagcccagc	aggctcatctg	ctccagccctg	1140
tcctctcgtc	agccttcctc	ttccagaagc	tggttgagag	acatttcagga	gagagcaagc	1200
cccttgatcat	gtttctgtct	ctgttcataat	cctaaagata	gacttctcct	gcaccgccag	1260
gaaagggtag	cacgtgcagc	tctcaccgca	gatggggcct	agaatcaggc	ttgcttgagg	1320
gcctgacagt	gatctgacat	ccactaagca	aattttattta	aattcatggg	aaatcacttc	1380
ctgccccaaa	ctgagacatt	gcattttgtg	agctcttggg	ctgatttgga	gaaaggactg	1440
ttacccattt	ttttggtgtg	tttatggaag	tgcatgtaga	gcgtcctgcc	ctttgaaatc	1500
agactgggtg	tgtgtcttcc	ctggacatca	ctgcctctcc	agggcattct	caggccccgg	1560
gggtctcctc	cctcaggcag	ctccagtggt	gggttctgaa	gggtgctttc	aaaacggggc	1620
acatctgggt	gggaagtcac	atggactctt	ccaggggagag	agaccagctg	aggcgctctc	1680
ctctgaggtt	gtgttggtc	taagcgggtg	tggtggggc	tccaaggagg	aggagcttgc	1740
tgggaaaaaga	caggagaagt	actgactcaa	ctgcactgac	catgttgta	taattagaat	1800
aaagaagaag	tggtcggaag	tgacacattc	tggtataggaa	tcacagctca	ccccaggatc	1860
tcacaggtag	tctcctgagt	agttgacggc	tagcggggag	ctagtctccg	cgcatagtta	1920
tagtggtgat	gtgtgaacgc	tgacctgtcc	tgtgtgctaa	gagctatgca	gcttagctga	1980
ggcgccctaga	ttactagatg	tgtgttatca	cggggaatga	gggtgggggtg	cttatttttt	2040
aatgaactaa	tcagagcctc	ttgagaaatt	gttactcatt	gaactggagc	atcaagacat	2100
ctcatgggaag	tggatacggg	gtgatttggt	gtccatgctt	ttcactctga	ggacatttaa	2160
tcggagaacc	tcctggggaa	ttttgtggga	gacacttggg	aacaaaacag	acaccctggg	2220
aatgcagttg	caagcacaga	tgtgtccacc	agtgtctctg	accaccctgg	tgtgactgct	2280
gactgccagc	gtggtacctc	ccatgctgca	ggcctccatc	taaatgagac	aacaaagcac	2340
aatgttcaat	gtttacaacc	aagacaactg	cgtgggtcca	aacactcctc	ttcctccagg	2400
tcatttgttt	tgcattttta	atgtctttat	tttttgtaat	gaaaaagcac	actaagctgc	2460
ccctggaatc	gggtgcagct	gaataggcac	ccaaaagtcc	gtgactaaat	ttcgtttgtc	2520
tttttgatag	caaattatgt	taagagacag	tgatggctag	ggctcaacaa	ttttgtattc	2580
ccatgtttgt	gtgagacaga	gtttgttttc	ccttgaactt	ggttagaatt	gtgctactgt	2640
gaacgctgat	cctgcatatg	gaagtccrc	ttcggtgaca	tttcttggtc	attcttgttt	2700
ccattgtgtg	gatggtgggt	tgtgcccact	tcctggagtg	agacagctcc	tggtgtgtag	2760
aattcccggg	gcgtccgtgg	ttcagagtaa	acttgaagca	gatctgtgca	tgcttttctc	2820
ctgcaacaat	tggtctgttt	ctcttttttg	ttctcttttg	ataggatcct	gtttcctatg	2880
tgtgcaaaat	aaaaataaat	ttgggcacaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2940
aaaaaaaaag						2950

<210> 93

<211> 1722

<212> DNA

<213> Homo sapiens

<400> 93

ggcacgagcc	agagcaggct	gctaggccctg	gggccaccac	tgccccctggg	tgctacaccc	60
agtgtgctgg	gtcactggga	acttcctgaa	tggtgtctac	ctgaactggg	cccccaagga	120
tggtgtgctg	gcagtaccgc	aggaagagga	gcagcccctg	tgaagattga	gagctgccag	180
aggctctgtg	attggctgcg	gcacgatgac	ccgcgcacgg	attggctgct	tcggggccgg	240
gggccggggc	cgggggacag	aatccgcccc	cgaaccttca	aagagggtac	cccccggcag	300
gagctggcag	accyaggagg	tgacacagac	ccgcggggca	aacggactgg	ggccaagagc	360
cgggagcgcg	ggcgcaaaag	caccagggcc	cgcccagggc	gcccgcgacg	acggccttgg	420
gggttctcgc	ggccttcggg	tgcgctctc	gcctctagcc	atggggctcg	cagcgttgga	480
gatccctggg	ctggtgctgt	gcctgggtgg	ctgggggggt	ctgacctctg	cgtgctgggt	540
gcccattctg	caggtgaccg	ccttcctgga	ccacaacatc	gtgacggcgc	agaccacctg	600
gaaggggctg	tggatgtcgt	gcgtgggtgca	gagcacsggg	cacatgcagt	gcaaagtgtg	660
cgactcgggt	ctggctctga	gcaccgaggt	gcaggcgggc	cgggcgctca	ccgtgagcgc	720
cgtgctgctg	gcgttcgttg	cgctcttctg	gaccctggcg	ggcgcgagct	gcaccacctg	780
cgtggccccg	ggcccggcca	argcgctgt	ggccctcaag	ggaggcgtgc	tctacctgtt	840
ttgggggctg	ctggcgctcg	tgccactctg	ctgggttcgcc	aacattgtcg	tcgcgaggtt	900
ttacgacccg	tctgtgcccc	tgctgcagaa	gtacgagctg	ggcgcgagcg	tgtacatcgt	960
ctgggcggcc	accgcgctgc	tcattggtagg	cggctgcctc	ttgtgctgcg	gcgcctgggt	1020
ctgcaccggc	cgtcccagac	tcagcttccc	cgtgaagtac	tcagcgccgc	ggcgggccac	1080
ggccaccggc	gactacgaca	agaagaacta	cgtctgaggg	cgctgggcac	ggccggggcc	1140

ctcctgccag	ccacgcctgc	gaggcgcttg	ataagcctgg	ggagccccgc	atggaccgcg	1200
gcttccgcg	ggtagcgcg	cgcgagggct	cctcggaacg	tccggctctg	cgccccgacg	1260
cggctccttg	atccgctcct	gcctgcgccc	gcagctgacc	ttctcctgcc	actagcccg	1320
ccctgcctt	aacagacgga	atgaagtttc	ctttctctg	cgcgcgctg	tttccatagg	1380
cagagcggt	gtcagactga	ggatttcgct	tccctccaa	gacgctggg	gtcttggtg	1440
ctgccttact	tcccagaggc	tcctgctgac	ttcgagggg	cggatgcaga	gcccagggcc	1500
cccaccggaa	gatgtgtaca	gctggctctt	actccatcgg	cagggcccga	gcccagggac	1560
cagtgaattg	gcctggacct	cccggctctc	ctccagcatc	tcccaggca	aggcttggtg	1620
gcaccggagc	ttgagagagg	gcgggagtg	gaaggctaag	aatctgctta	gtaaatgggt	1680
tgaactctca	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aa		1722

<210> 94

<211> 635

<212> DNA

<213> Homo sapiens

<400> 94

gcctaaagag	agctccccca	ggaccagccc	tgccaagg	attgctgcag	ccctcatcca	60
ccttccaagc	actggaaca	aacattggag	accaagtga	gcgtcactca	acagccgtag	120
taatcaggga	aatgacaagt	tacatactga	tatcctttgt	tttgcgtgatt	ggagttgggt	180
gcattgaaaa	agatcagtcg	tgcccagtg	tggggggaag	gaagcgtctt	cacctgttgt	240
ttgtgggagg	acagttgagg	caggtgagga	tgctgagagg	tgagctcagc	tgtgctgtt	300
accgtccaca	tgtgcaagcc	cttcagctcg	gtggtgttac	ttgttttga	gatgcagttt	360
cactcttgtc	accaggctg	gagtgcattg	catgatcttg	gctcgtgca	acatccgct	420
cccgggttca	agcgattctc	ctgtctctac	taaaaataca	aaaattagct	gggtgtggtg	480
gtgcgtgcct	ttaatcccag	ctactcagaa	ggctgagggtg	caagaattgc	ttgaacctgg	540
gaggtggagg	ttgccgtggg	ccgagatcac	gccaccgcac	tccagcctag	gcaacagagc	600
tagactgtct	caaaaaaaaa	aaaatgacct	tcgag			635

<210> 95

<211> 3798

<212> DNA

<213> Homo sapiens

<400> 95

ccacgcgtcc	ggggcttcat	acaggaaatc	tattgctgtg	tcaagttcca	gagaaaaagc	60
tctgttcgtc	caagttacta	accaggctaa	accacataga	cgtgaaggaa	ggggctagaa	120
ggaagggagt	gccccactgt	tgatggggt	agaggatcct	gtactgagaa	gttgaccaga	180
gaggggtcca	ccatgcgcac	agttcctctt	gtaccagtg	ggaggaaaag	tactgagtga	240
agggcagaaa	aagagaaaac	agaaatgctc	tgccttggga	gaactgctaa	cctagggcta	300
ctgttgattt	tgactatctt	cttagtgccc	gaaggggagg	gtgctgctca	accaaacaac	360
tcattaatgc	tgcaactag	caaggagaat	catgctttag	ctccaagcag	tttatgtatg	420
gatgaaaaac	agattacaca	gaactactcg	aaagtactcg	cagaagttaa	cacttcatgg	480
cctgtaaaaga	tggctacaaa	tgtgtgtctt	tgttgccctc	ctatcgcat	aagaaatttg	540
atcataataa	catgggaaat	aatcctgaga	ggccagcctt	cctgcacaaa	agcctacaag	600
aaagaaacaa	atgagacca	ggaaaccaac	tgtactgatg	agagaataac	ctgggtctcc	660
agacctgatc	agaattcgga	ccttcagatt	cgtaccgtgg	ccatcactca	tgacgggtat	720
tacagatgca	taattggaac	acctgatggg	aatttccatc	gtgcatatca	cctccaagtg	780
ttagtctacac	ctgaagtgc	cctgtttcaa	aacaggaaata	gaactgcat	atgcaaggca	840
gttgccaggga	agccagctgc	gcatactctc	tggtatccag	agggcgattg	tgccactaag	900
caagaatact	ggagcaatgg	cacagtgaat	gttaagagta	catgccactg	ggaggtccac	960
aatgtgtcta	ccgtgaactg	ccacgtctcc	catctgactg	gcaacaagag	tctgtacata	1020
gagctacttc	ctgttccagg	tgccaaaaaa	tcagcaaaat	tatatattcc	atatatcatc	1080
cttactatta	ttattttgac	catcgtggga	ttcattttgg	tgttgaaagt	caatggctgc	1140
agaaaatata	aactgaataa	aacagaatct	actccagttg	ttgaggaggga	tgaaatgcag	1200
ccctatgcca	gctacacaga	gaagaacaa	cctctctatg	atactacaaa	caaggtgaag	1260
gcactctgagg	cattacaaa	tgaagttgac	acagacctcc	atactttata	agttgttggg	1320

ctctagtacc	aagaaacaac	aacaaacgag	atacattata	attactgtct	gattttctta	1380
cagttctaga	atgaagactt	atattgaaat	taggttttcc	aaggttctta	gaagacattt	1440
taatggattc	tcattcatac	ccttgataaa	ttggaatttt	tgattcttag	ctgctaccag	1500
ctagttctct	gaagaactga	tgttattaca	aagaaaatac	atgcccata	ccaaatattc	1560
aaattgtgca	ggacagtaaa	taatgaaaac	caaatttctt	caagaaataa	ctgaagaagg	1620
agcaagtgtg	aacagtttct	tgtgtatcct	ttcagaatat	tttaatgtac	atatgacatg	1680
tgtatatgcc	tatggatat	gtgtcaattt	atgtgtcccc	ttacatatat	atgcacatat	1740
ctttgtcaag	gcaccagtgg	gaacaatata	ctgcattact	gttctataca	tatgaaaacc	1800
taataatata	agtcttagag	atcattttat	atcatgacaa	gtagagctac	ctcattcttt	1860
ttaatgggta	tataaaattc	cattgtatag	ttatatcatt	atttaattaa	aaacaaccct	1920
aatgatggat	atttagattc	ttttaagttt	tgtttatttc	ttttaagttt	tgtttgtggg	1980
ataaacaata	ccacatagaa	tgtttcttgt	gcataatat	ctttgttttt	gagtatatct	2040
gtaggataac	tttcttgagt	ggaattgtca	gggtcaaagg	tttgtgcatt	ttactattga	2100
tatatatggt	aaattgtgtc	aaatatatat	gtcaaattcc	ctccaacatt	gtttaaatgt	2160
gcctttccct	aaatttctat	tttaataact	gtactattcc	tgcttctaca	gttgccactt	2220
tctcttttta	atcaaccaga	ttaaatatga	tgtgagatta	taataagaat	tatactattt	2280
aataaaaaatg	gatttatatt	tttggtcatg	tttctaagag	agtgaatgca	cgtgtgagaa	2340
catttagcttc	ttctgaactc	attatatctc	cacagagggtg	ttgatacttg	atgcctaaca	2400
gttttgcaga	tgtgtcatat	tggaattgtg	tatttttatg	gtgtacattc	tattgtgata	2460
tatttatgta	ataattaatg	tctattgacc	atataagtg	cgaaaaatgc	accatagagg	2520
acatggggta	tttattttaca	aactatgagc	tacataataa	gcaagtggcc	atgggatggc	2580
atgaccctcc	cctccatatt	tttgtggagc	aaaatatggg	caatgtttat	gtaaatcatt	2640
gttaatatca	tgaattattt	tttaattaaa	aacataagtc	tatttgcctc	atagcagaaa	2700
aaacatgaga	agttttttca	tcatgataga	aattgaaaca	aaatatattc	attcttcaat	2760
cataccatct	gagattttta	agacagctat	tttgtcttat	aagtataatt	ttctccctct	2820
agacatttca	gttactatgg	attttgtcct	caaagggacc	tttagtctat	tttggatgta	2880
aagctaattc	aatgacactt	ggcacatgat	attttgatca	agccattttg	acttgaccaa	2940
aaagcagtg	ccattaggtt	tctgcatata	aatattacca	agcaatgttc	acaatagaca	3000
tcattacact	gtccttgaaa	tttatttaatt	tttcatccaa	ccctgggttg	gctgaggctc	3060
atagtttagt	tcaagactat	ctgtttaaat	attactgaaa	aacaaagtaa	gacagtacta	3120
tgcttacctc	tttaacttgat	aatgtcaaaa	caggcatgtt	aatgacatc	atagaaaaga	3180
cttcaagata	atttatagaa	gttaaatat	attgtacaga	aaataattgt	atgaaaatct	3240
ctactatggg	gctggaacat	gggtgaacat	tagaatgata	taaaaaatta	tatatattct	3300
ccaaatccac	gctagacctg	tcaaattaga	gaatctagag	attagacctg	gcgtgtcagc	3360
aagggtcatc	aggaagcaga	ggctgagacg	gagttagggtg	tgattactta	catagtcgat	3420
tacattttac	aaataacatt	ttatatgtct	catttactgt	gctttctccc	catcccattt	3480
tgtatctttt	cctttgcttt	gctagatttg	tcaattttct	ctctctttct	gtctctctct	3540
ctttcaatat	ctctaataat	ttgaaagtaa	ttcatcataa	ctaaatatct	attgggggta	3600
tgtctcactc	acaaacttct	gaaaacggct	ttactgagat	ataattgata	tatttaagt	3660
tacagtttgt	taaaatttgc	acatatttta	aatgtggact	ttggtaaatg	ttgacatagt	3720
tttacatctg	tgaaccatc	agcataatca	agataataaa	cttgtccatc	accccccaaa	3780
aaaaaaaaaa	aaaaaaaaaa					3798

<210> 96

<211> 2683

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (2640)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (2676)

<223> n equals a,t,g, or c

<400> 96

acaasgtmac	gcctgacag	tmacccggatc	cggaattcc	cggtcgacc	caaggcgctcc	60
gcatttgcaa	taacagaaaa	ggaattgcat	gtatgaagtt	ttcaatcgtg	ggctttttctt	120
tggttgggg	aggggggtcg	gggatagttt	gatttccatt	ttctgaaaac	gacagacttg	180
gattctgttt	gtgtgtgcat	attttatcca	gccttaagtt	ataaagctca	tctgtcccg	240
tgcattccct	gtgtattttc	aggacatggc	tcgtgggtgt	gtgtgttcat	tgtgtgcgtc	300
tgatgtatt	tttctgtcat	cactgttccc	tctctcccg	agtgtgcatt	cagttaatat	360
aatcagttgc	ttgcttcttt	caaagtgcct	tgaaggctct	gaactcatgt	gtgagcatct	420
ttatcaacta	tccaattgc	atgttctcca	tcacatatcc	tcttatttgc	tctgtacccc	480
ctgagaatat	gttttagaga	tattggaata	aagctgtctg	ggtaaggagt	aggcttagcc	540
gacctatgaa	taatacactt	tagtctagtt	ctttatttcta	aatctggatt	gccagtattg	600
tgtatttaaa	ccaagtctgt	gaatacctgc	tttttttggc	cacagagtaa	caagttttca	660
tgtaagatct	tcataccaaa	gtaggaagta	aaaatagctt	agaaagctct	gtcaggtgtt	720
ttgtgcagct	gacagargta	atgttacatc	acctaaaaaa	gaaagatata	cggtcagtta	780
tcctaaaaat	aaattgtttg	gaaagtacaa	tgcaccacat	ttttgtagaa	gtctactatt	840
tgataaacag	ttgaaaattc	agatgtgttt	gaccttagt	catttttact	ctttggttct	900
gagtatacct	attttcttag	cgatatctgc	ttgtttatct	ttttcttcac	cttttaacaa	960
gtatgacata	ggaaagtcac	tttttttttag	aattcatgga	tcagtctgat	ctactcttat	1020
tcataatgga	acatgtaaat	atactgaaaa	ctgttttttca	ggagagaaat	atgagttgga	1080
gggaaggaaa	agtggttcta	ctaatgttcc	aaaatcctca	tcagagaagg	tatgatgttc	1140
tcagggtgtg	aaaatatttt	ttagttgatt	gagaatgcag	gtttaacaga	agagataagg	1200
ggcataatga	ctgctgggtt	tccagactgg	attttccctac	cgcaactatt	aatgttctca	1260
gagttgatga	ggaccacctt	tgtgtataca	cttgtagttt	taaaccttgc	artggtaaca	1320
aaatgatcaa	ctttaatcca	ggtagaattc	aagatggctg	tacttcagtt	gtatgataaa	1380
attaatggtt	ctcatgactt	gtgtggcatc	taaaaaataat	gtttttatag	catctctctg	1440
ccactaaatt	gttgacttga	attttgggaa	aaaaaaaagt	tggtgttgat	atgtatatgt	1500
gtgtgtgtat	atatgtattt	ataaacaagt	gtgtttgagt	aacaagttag	tttcatagtc	1560
ttcccctacg	catgtgtatt	ccacacacaa	atggctgagt	tatagtcata	aaacaatttg	1620
caataaaaaa	aaaaccacaa	cagattgtca	gttaaccagg	aaacagttaa	tgtttttttaa	1680
tgaatctggc	attatagtga	gcaaattgtc	tattaattta	ggctaatttc	taatactacc	1740
ataatttgtg	tctaaatttc	tggtggggta	gaaattacta	aaattgtggg	gagttttttc	1800
tgatttttac	attgcttttag	gaaacatttt	tactaattca	gctgtcttag	gtaaaaatgaa	1860
tagttttctt	cctgtttttt	tatgtgtcat	tgttagtggt	ctcagaattc	tgatcagtaa	1920
ctttgtgtat	gatgtgaat	tacaaaccgt	ttgaatgatc	cagttgaaaa	cgtatccctc	1980
tactttcttc	agttgtagaa	aaggtttaatt	tccctcagtg	tcccacatta	taccaacctta	2040
agagaagaac	aggtaatagg	gagaaaataaa	catacgggtg	tttcagtggt	tttgggtcatg	2100
tgtccacagg	agaaactaac	cattcagttg	tcttaatttt	agttcggttc	accctgtgag	2160
gagtttgttt	ccatcagttg	ttgactttcc	aaaatgttgc	attaagtaat	agttgtcact	2220
ctgttgggtc	catgggtcaat	atcaatcaga	ctttcatgat	ctctactaat	tattagtaga	2280
gtcctgtact	atgtctgtaa	ctactaagtt	taaagaaaag	cacatagtca	cttcatctct	2340
ttttttctta	gcctacgtct	actccccaac	ccatcccaac	attgacatgc	tatctgtgga	2400
caaatagcag	ttctcagaat	ctagtcaagt	tgccatcatc	ccccctgcct	tggccgttca	2460
tagtaggtat	gcataatgtt	gtttctgtac	agtactgtgt	gtgtgtgtgt	atatatatat	2520
acatctgtat	gcacacatct	ttgataaaaat	agctatttga	ctagcagggt	taaagtggct	2580
tttaattact	tcgtgagtg	tattggatac	atcttaaaaa	aaaaaaatct	ggaccacagan	2640
ccatgccata	cttgggttga	ctattttggg	gcattnaaaa	ttg		2683

<210> 97

<211> 2181

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (5)

<223> n equals a,t,g, or c

<400> 97

gtacnggatt	cccgggtcga	cccacgctc	cggcaggcag	aggggcagtg	ggcgcgggga	60
agatgattct	gggcctcccc	catctactgt	cattaacca	aatgaaacat	ttgccaaacat	120
aatttttaaa	cctactgtag	tacaacaagc	caggattgcc	cagaatggaa	ttttgggaga	180
ctttatcatt	agatatgacg	tcaatagaga	acagagcatt	ggggacatcc	aggttctaaa	240
tggctatttt	gtgcactact	ttgctcctaa	agaccttcct	cctttaccca	agaatgtggt	300
attcgtgctt	gacagcagtg	cttctatggt	gggaacaaaa	ctccggcaga	ccaaggatgc	360
cctcttcaca	attctccatg	acctccgacc	ccaggaccgt	ttcagtatca	ttggattttc	420
caaccggatc	aaagtatgga	aggaccactt	gatatcagtc	actccagaca	gcatcaggga	480
tgggaaagtg	tacattcacc	atatgtcacc	cactggaggc	acagacatca	acggggctct	540
gcagagggcc	atcaggctcc	tcaacaagta	cgtggccccc	agtggcattg	gagaccggag	600
cgtgtccctc	atcgtcttcc	tgacggatgg	gaagcccacg	gtcggggaga	cgcacaccct	660
caagatcctc	aacaacaccc	gagaggccgc	ccgaggccaa	gtctgcatct	tcaccattgg	720
catcggcaac	gacgtggact	tcaggctgct	ggagaaactg	tcgctggaga	actgcggcct	780
cacacggcgc	gtgcacgagg	aggaggacgc	aggctcgcag	ctcatcggtt	tctacgatga	840
aatcaggacc	ccgctcctct	ctgacatccg	catcgattat	ccccccagct	cagtgggtgca	900
ggccaccaag	accctgttcc	ccaactactt	caacggctcg	gagatcatca	ttgcggggaa	960
gctggtggac	aggaagctgg	atcacctgca	cgtggaggtc	accgccagca	acagtaagaa	1020
attcatcatc	ctgaagacag	atgtgcctgt	gcggcctcag	aaggcaggga	aagatgtcac	1080
aggaagcccc	aggcctggag	gcgatggaga	gggggacmcc	aaccacatcg	agcgtctctg	1140
gagctacctc	accacaaagg	agctgctgag	ctcctggctg	caaagtgcag	atgaaccgga	1200
gaaggagcgg	ctgcggcagc	gggcccaggc	cctggctgtg	agctaccgct	tcctcactcc	1260
cttcacctcc	atgaagctga	gggggcccgt	cccacgcgat	gacggcctgg	aggaggccca	1320
cggcatgtcg	gctgccatgg	gaccggaacc	ggtgggtgcag	agcgtgcgag	gagctggcac	1380
gcagccagga	cctttgctca	agaagccata	ccagccaaga	attaaaatct	ctaaaacatc	1440
agtggatggt	gatccccact	ttgttggtga	tttccccctg	agcagactca	ccgtgtgctt	1500
caacattgat	gggcagcccg	gggacatcct	caggctggct	tctgatcaca	gggactctgg	1560
tgtcacagtg	aacggagagt	taattggggc	acccgccccct	ccaaatggcc	acaagaaaca	1620
gcgcacttac	ttgcgcaacta	tcaccatcct	catcaacaag	ccagagagat	cttatctcga	1680
gatcacaccg	agcagagtca	tcttggtatg	tggggacaga	ctggtgctcc	cctgcaacca	1740
gagtggtgtg	gtggggagct	ggggkctgga	ggtgtccgtg	tctgccaaag	ccaatgtcac	1800
cgtcaccatc	cagggtctcca	tagcctttgt	catctctatc	cacctctaca	aaaagccggc	1860
gcccttccag	cgacaccacc	tgggttttcta	cattgccaac	agcaggggcc	tttccagcaa	1920
ctgccacgga	ctgctgggtc	agttcctgaa	tcaggatgcc	agactcacag	aagaccctgc	1980
agggccccagc	cagaacctca	ctcaccctct	gtcctctcag	gtgggagagg	ggcctgaggc	2040
cgtcctaaca	gtgaaaggcc	accaagtccc	agtgtgtctg	aagcaaagga	agatttataa	2100
cggggaagag	cagwtagayt	gytggtttgc	caggaaacatg	ccgccaaact	gattgacggg	2160
gagtacagga	ttacctggca	t				2181

<210> 98

<211> 1957

<212> DNA

<213> Homo sapiens

<400> 98

acgcgtccgg	aggcggcagg	ctcatggcgc	cggcgctcgcg	gttgctcgcg	ctctggggcgc	60
tggcggctgt	ggctctaccc	ggctccgggg	cggagggcga	cggcgggttg	cgcccgggcg	120
ggccgggggc	cgtggcggag	gaggagcgct	gcacgggtga	gcgtcggggc	gacctcacct	180
acgcggagtt	cgtgcagcag	tacgccttcg	tcaggcccg	catcctgcag	ggactcacgg	240
acaactcgag	gttccggggc	ctgtgctccc	gcgacagggt	gctggcttcg	tttggggaca	300
gagtgggtccg	gctgagcacc	gccaacacct	actcctacca	caaagtggac	ttgcccttcc	360
aggagtatgt	ggagcagctg	ctgcaccccc	aggaccccac	ctccctgggc	aatgacaccc	420
tgtactttct	cggggacaac	aacttcaccg	agtgggcctc	tctctttcgg	cactactccc	480
cacccccatt	tggcctgctg	ggaaccgctc	cagcttacag	ctttggaatc	gcaggagctg	540
gctcgggggt	gcccttccac	tggcatggac	ccgggtactc	agaagtgate	tacggctgta	600
agcgtgggtt	cctttaccca	cctgagaaga	cgccagagtt	ccaccccaac	aagaccacgc	660
tggcctggct	ccgggacaca	tacccagcct	gcaccgtctg	cacggccctg	gagtgatcca	720
tccgggctgg	tgaggtgctg	acttcccgc	cgctgggtgg	atgctacgct	caaccttgac	780
accagcgtct	tcattctacc	ttcctcggct	agccaaaaca	gctggcagga	ctgccgggtca	840

cacaccagca	cgccccacct	cgtgctcacg	gattttatta	cacagatagt	ggcggcaatg	900
gcctcagccc	agccccacct	cacctgcttt	tccagccac	aaagggggac	gatcacggcc	960
cagcaaaagc	gatgctgaga	gggaaacag	tccagagtcc	aacagcagaa	cttgggggaa	1020
gcggtcgggg	tggccaggaa	cataaactat	gtataggggc	cgggggcttc	tgcccagggc	1080
tcccctggac	caggacgcca	ggtagggcag	ggaacctcag	tagtctctca	cccagccatt	1140
ctcagagatg	aatgcgtcaa	taacctcctt	catagccaag	ttggggatga	gctgttcctg	1200
ggtcaggggg	ctccgggtca	cgggggtcaaa	atgaccacac	cgctgcagtg	acaagaaggg	1260
cagagggcag	tcatggggcc	caggaccatg	ccactggccc	tgctccccc	gccgcaggcc	1320
tcacctgcag	gtgctcctcg	atgtccttgc	ggtcgtaggt	gatgccactg	ggcgtgatgc	1380
acgggtcccc	catcagctca	aagctgatct	tgccacacag	gtagtcgggg	atgtctcgct	1440
tctgtggcac	aggggcacac	ggtcagaggc	tgaaggggg	cactgcacga	gcacctgcca	1500
gccatcggca	gcaagcgaca	cacactcacc	ttcctcttct	catccacctg	agaaaaaagc	1560
tcgtccatgt	ccgcatgta	cttgtcctgt	gaagagttga	gtgctgtgct	tgggggagac	1620
acccacacct	cctcctccat	ggggcacaga	cccaacacaa	ggcggggatg	ctcccacgcc	1680
acgtgcacac	acacagaccc	acatgtgggt	ggggggcacc	ctcacgtgct	tgccctcaat	1740
gcaggcctgc	tggggccgga	cgtggctgtc	gtcctcatca	ccctcggtgt	ttcgctggca	1800
ctcttcagc	tccctggggg	ttgaccagga	gcccgtcaga	gatggacctg	gccagatgtc	1860
tgaccacacc	ccaatctcag	agctaatac	cacacttccc	cacatttccct	gcttgccagt	1920
aaagccttcg	ataaacaaaa	aaaaaaaaaa	aaaaaaa			1957

<210> 99

<211> 1112

<212> DNA

<213> Homo sapiens

<400> 99

ccacgcgtcc	gcccaccaac	aaaggcactt	gcagttacat	tcaccacatt	tgtaacggag	60
ccattgaagc	atattggaaa	aggaactggg	gaattttatta	aagcactcat	gaaggaaatt	120
ccagcgctgc	ttcatcttcc	agtgtgata	attatggcat	tagccatcct	gagtttctgc	180
tatgggtgctg	gaaaatcagt	tcatgtgctg	agacatatag	gcggctctga	gagagaacct	240
ccccaggcac	ttcgggccag	ggatagaaga	cggcaggagg	aaattgatta	tagacctgat	300
ggtaggagcag	gtgatgccga	tttccattat	agggggccaaa	tggggccccc	tgagcaaggc	360
ccttatgccca	aaacgtatga	gggtagaaga	gagattttga	gagagagaga	tggtgacttg	420
agatttcaga	ctggcaacaa	gagccctgaa	gtgctccggg	catttgatgt	accagacgca	480
gaggcacgag	agcatccac	ggtaggtacc	agtcataaat	cacctgtttt	ggatacaaa	540
cccaaggaga	caggtggaat	cctgggggaa	ggcacaccga	aagaaagcag	tactgaaagc	600
agccagtcgg	ccaagcctgt	ctctggccaa	gacacatcag	ggaatacaga	aggttcaccc	660
gcagcggaaa	aggcccagct	caagtctgaa	gccgcaggca	gcccagacca	aggcagcaca	720
tacagccccg	caagaggtgt	ggctggacca	cgtggacagg	atccgggtcag	cagcccctgt	780
ggctagagga	acaccagcac	aaacgacagc	ctcaagctct	cttcgagctt	tgtaaaagtaa	840
tctcacaagc	tgccaaagaa	cacatcccaa	gcagtaatca	cagatgatga	agcaccttac	900
aggacccctc	caccctcaaa	cgcgcagtgt	cagagaagta	ttaatgcctt	aatagactac	960
tgaagggttaa	ctattttacat	catttctaaaa	ataattttaag	ctgtattaca	gtgctataaa	1020
ttctcctttt	aaggaaaaaa	gtaatatagatt	gtttcctttag	cgtgtttgaa	aggaaaaaaa	1080
ataaatccat	agtaactaaa	aaaaaaaaaa	aa			1112

<210> 100

<211> 887

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (303)

<223> n equals a,t,g, or c

<400> 100

ggcacgagag aattataggt gcatgatggg gcttttggag actggcaatg ttctgttttg	60
ggtctgggta gtggctactt gtgtgtattc actttatgct aactcattga actgtacaga	120
tatggactgt gcccctttct atatgtgtgt aatgcttcaa caaaagtgtc aatagtgatg	180
catgcacatg ttaaaatttc aaactatatt aaagagtgtg caattaaaag gaagttatcc	240
ttctactcta aagtcctatt tctctcttc agtctatcat tactactagt ttctagtata	300
tcnttttagaa atgkgctgta aaagaacaat gatgtgtatg tctctacagg tatatatata	360
ttctttttgt aattacaaaa atataaactc actacatatg ttattctacc acatgctttt	420
tttcatgtaa cagtatgtct tagacatctt ttcttattat tgcgtggaag ctatcaacct	480
tattctttgt aatgactcca taaattattc tactgtgaaa acacaccatg ttaaccagt	540
tctctgttag tgaacattta ggatttttcc agttttttaa tattacagtg acatattaaa	600
cattaaacat atcttttgac acatgtcctt gcacacatgt ataggatga tggactttaa	660
caccctttgg ctagattctt tagcacataa cgtaaatatc ccatagagtc aaaaccacc	720
ttaaaacttc ctacaggaggc tgggtgtggt ggctcacgcc cgtaatccca gcactttagg	780
aggccgaggt gggcgatca cgaggtcagg agatcgagac catcctggcc aacatggtga	840
aaccctgtct ctactaaaaa taaaaaaaaa aaaaaaaaaa actcgta	887

<210> 101

<211> 1248

<212> DNA

<213> Homo sapiens

<400> 101

ccacgcgtcc ggtctattaa ggtgacagaa ctgaagggcc tggccaacca tgtggttgtg	60
ggcagtgtct cctgtgagac caaggacctg tttgccgccc tgccccaggt tgtggctgtg	120
gatatcaatg accttggtac catcaagctc agcctggaag tcacatggag ccccttcgac	180
aaggatgacc agccctcagc tgcttcttct gtcaacaagg cctccacagt caccaagcgc	240
ttctccacct atagccagag cccaccggac acaccctcac ttcggaaca ggctttctat	300
aacatgctgc gacggcagga ggagctggag aatgggacag catggtccct gtcattgaa	360
tcttcagacg actcatccag cccacagctc tcaggcaactg cccgccactc accagccct	420
aggccctggt tgcagcagcc cgagccctt cccatccaag ttgccttcg caggcctgag	480
acccccagct ctgggccctt ggatgaggag gggggcgtgg cccagtcct ggcaaatggg	540
catgcaccct acagtgggac tctgagccac atcagtggag ctagtgtaaa tgcctgcttg	600
gctgaggctt cagtggaggc cgttgggcca aaaagcctat cctggggacc tagccacct	660
acacaccag ctcccaccca tggaaagcac cccagtcctg ttctctctgc cctggaccct	720
ggccactctg ccacaagctc taccctcggg acaacaggct ctgtcccccac atctacagac	780
cctgcccctat ctgcacacct agactcagtt cataagtcca cagactctgg ccttcagaa	840
ctgccaggcc cactcacac cactacaggt tctacctata gtcctattac cactaccac	900
agtgtccaa gccccctcac tcacactact acaggctcca cccacaagcc cataatctct	960
acccttacta ctacaggccc taccctcaat atcataggcc cagtccagac taccacaagc	1020
cccaccaca ctatgccaag ccttctctcc cacagcaata gtccccaata cgtagatttt	1080
tgctctctg tatgtgaaa catttttcta cattatgta ttggaatttt ctctcatata	1140
ttatattcct ctaaaactct caaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1200
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	1248

<210> 102

<211> 1841

<212> DNA

<213> Homo sapiens

<400> 102

ggcacgagct ggtgcaggag ctggagcagt accagttgtt gccgaagaga ttggactggg	60
agggaacga gcacaacagg agctacgagg agttggctct gtccaataag catgtggctc	120
ctgatcatct tttgcaaatc tgccagcgca tcggctctat gttggataaa gaaattccac	180
ccagtatttc aagagtcact tctttacttg gtgcaggaag gcagtcttg ctacgtacag	240
caaaaggtag ctttaattga agaagtgtc tgccttgtat taatagccta ataattatta	300
taattgtgaa gtaatgtaag catgtggaag aaaggatga atgtaactat aaatcatgct	360
gtggactata aaattctctg catttagcat agaacttga attttattac ataagtgttt	420

aatgggtcct	tctgcattaa	aaaagttgat	aaagatatag	atgttaagct	aaaaagcttg	480
aattttattct	gagatgttgt	ttgagggtttt	atcaatgtga	gttcttaatt	ttccattttg	540
tagctcactc	ataacttcat	tgaatgtaat	tgtatgtgtc	taataaaactc	tttaaatgct	600
agaggatattt	catttgatat	ttagcttctg	aaagcccttt	gttataaagg	gaagtgtattt	660
tctaaattttg	gtagcagctt	gaaaacaagc	accacaaatg	taaaattgca	tgggtctttt	720
tttactaatt	taattttttt	gccaaagtta	gtgctagcaa	aatgtttgaa	attgctattc	780
tttttctgtt	agcttggtat	atatgactta	gactaagtat	atgtatgtga	tataatatat	840
gtgtacatgt	ggctaaactg	ctaattgggca	aagattgttg	gttcttctga	tgctgagtat	900
acatcaagtt	catcccatga	ttatttttta	ttacgagcta	ttgctgatgt	gatgctactc	960
ttagctgcac	caccagtata	aagcttaact	gagaaagaac	acccctccct	cttagttccg	1020
tgccagactg	gtttgttagc	tttagttgtt	tcttagtgct	ctagccatgt	caccctgtgg	1080
cactaatcag	gacactgttt	atattcttcc	tcttctgtgt	ccctattcac	tttgtaatta	1140
ttctattttta	gctgggtata	tggtagttaa	ttggtagcct	gttttctttt	gttttcttta	1200
cttggtgtcaa	gtcccatgtc	gttgttttgt	gatgagcagt	gatttagaac	tggtccagacc	1260
ttagaagagt	cttcagcaag	ctgccctagg	gtaggggcat	attctctgtt	tgttttgtac	1320
tggtacccca	ggaggatgcc	ttgtaactgc	tctttggcca	tgatagtggg	aatgaatcca	1380
ttcatgtaaa	atgcaatatg	tctgttagtt	tggggcaccc	gaggctcagt	agaggtttag	1440
tatgatgcc	tccctttctt	ctttgagtta	tgaagaaact	aatttctggc	acagtgtgac	1500
tgcttcaaac	ctatgtttctg	ctgctatctc	ctatagagga	gggcagactt	ttttttctta	1560
gtgcatttgc	tggtaggtca	aggagggatg	gactttgaat	tagtggttcag	agtcacagat	1620
cctacgtatg	tgctcttcag	tagaggattt	tctgtgatcc	tacaatgaag	ggaaagctat	1680
ataattttatt	ggtacttcct	aaagatatag	actaaagtca	tggtatatt	ggccaaatta	1740
atttttaaacc	tgaacaagtc	atgaattttc	atataaatca	aagaactttt	tccttataaa	1800
agcttccatt	tttgaaaagt	tgaaaaaaa	aaaaaaaaa	a		1841

<210> 103
 <211> 685
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (678)
 <223> n equals a,t,g, or c
 .
 <220>
 <221> SITE
 <222> (679)
 <223> n equals a,t,g, or c

<400> 103						
actacggctg	cgagaagacg	acagaagggg	ggcggcgacg	gaggaggagg	atggaggcgg	60
tggtgttcgt	cttctctctc	ctcgatttgt	gcgcgctcat	cttctctctg	gtctacttca	120
taattacatt	gtctgattta	gaatgtgatt	acattaatgc	tagatcatgt	tgctcaaaat	180
taaacaagtg	ggtaattcca	gaattgattg	gccataccat	tgctactgta	ttactgctca	240
tgtcattgca	ctggttcata	ttccttctca	acttacctgt	tgccacttgg	aatatatatc	300
gatacattat	ggtgccagat	ggtaacatgg	gagtgtttga	tccaacagaa	atacacaatc	360
gagggcagct	gaagtcacac	atgaaaagaag	ccatgatcaa	gcttggtttc	cacttgctct	420
gcttcttcat	gtatctttat	agtatgatct	tagctttgat	aaatgactga	agctggagaa	480
gccgtgggtg	aagtcagcct	acactacagt	gcacagttga	ggagccagag	acttctttaa	540
tcacctcttag	aaccgtgacc	atagcagtat	atatttttct	cttggaacaa	aaaactattt	600
ttgctgtatt	tttaccatat	aaagtattta	aaaaacatga	aaaaaaaaa	aaaaaaaaa	660
aaaaaaaaa	aaaaaaaana	aaaaa				685

<210> 104
 <211> 1168
 <212> DNA

<213> Homo sapiens

<400> 104

ggcagcagcc	cccgcacct	ccctccctcc	cctcccttcc	ccggccccgg	ctctggcccc	60
ggcccatctg	ctgttgggtc	ttctgctagg	gaggatgtcg	ggttcgtcgc	tgcccagcgc	120
cctggccctc	tcgctgttgc	tggtctctgg	ctccctcctc	ccagggccag	gcgccgtca	180
gaacgtgaga	gtacaatctg	gacaggatca	gaagtagaga	atgaagtgtg	aagagaaagg	240
gaaagacaga	agaaaggctg	cagtagtaca	aggagaaaag	caggatgcaa	gaatgaggaa	300
tgaatctttg	tttgaggagc	atccggaaaa	atataagctg	tcagaaagag	taaatagacc	360
agggacctct	aaagtaaatt	cacacatcaa	agttaaaata	atgttgagaa	atcacctcct	420
gtgaaaaatt	ggcttagctt	tcagtatgcc	tctttaaaca	aaacattatc	attttataca	480
aattttttaa	atgttggttca	taatatggag	tttaatttta	agcaccttag	ataatgcatt	540
tgtagcttgg	tttggtgaaat	ttttatggat	tttttttttc	aaaattccta	attttagttg	600
gtaaggatta	acttcgggaa	gacaggaaac	ccctccagta	aaattaattg	gttataaatg	660
gttacatatt	ttaggcttat	atacataaca	aaacttctaa	ctgaatttaa	aagtgatctt	720
gtgtaaaaaa	attattctag	tgatattgat	gtctaaattt	aaaagagtga	atacacaagt	780
aaaatatatc	tgctttttaa	ggatctattc	agtttaggaa	ggagagtcaa	aatttgtatt	840
taatttcaat	ttattatatt	gtctgaaact	gaaagtaggc	ctaacttttg	tttgcctttg	900
tgtaggtaca	aaggcaaaca	tttatcaatc	aactcttatt	aatttgagat	tattttgacc	960
tgattgctca	gaacttttgt	ctatctgtat	agaaatgggg	ctttctaaat	atttaaagaa	1020
tttcctacgt	atgtaaaatt	acttgacac	agtaagttag	gggagacatt	taatatccct	1080
cagataacctg	ctgttttctt	gttggttttc	ttgtttttca	aataaataaa	ctgagtgtta	1140
tctgttcatt	aaaatagaat	gtagtcag				1168

<210> 105

<211> 1175

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (24)

<223> n equals a,t,g, or c

<400> 105

gcatggaaat	atatcctata	catnaagtgc	tgctctgagc	tgtacatwta	gccatattta	60
grattttatt	argcaggatt	caaggcctcc	taaaaatgga	tattcagata	tttaaaaaata	120
cactattaag	gaatgcaatg	taaaatcaca	ctaacatttg	catagcactt	tatagtttac	180
aaaatgcctt	cacatatacg	agctcattta	ttctctcttc	ctctctttct	attcatctat	240
caagggaatat	cttctataag	ccaggcatca	ggcttgacac	tgaagacaca	aaatgaaaaa	300
gatatacaag	tctctattct	taaagagttc	gtagtttaat	gaagtgaccc	aaaaagtaaa	360
cagttgcaat	ataacatgca	aaggagtgtg	atgaaataca	cctagagtac	tatgaaaatg	420
gcatagaggg	accaagcctg	taatgaagtc	agagaaagta	aaagaggtga	agcckgagct	480
aagtttttaa	ggatgaacaa	aaattgacca	gacaaaggag	ttggagaagt	gggatgggtac	540
aggataaaaag	ccttttcta	ttgggggaagt	gtgagtcgtc	gatattagat	gatattatta	600
atagattgtg	cacagagtgc	acatggtaga	gaggtgagag	atgaggtagc	actgttaggg	660
cagataggtg	aattctcaca	gtaaccctgt	gwtgtacmtt	atcatcctcc	tctccattat	720
gtaggggaga	acaccaaaag	ccaaagagat	taagtaactt	tccagaattt	tgattttcaa	780
tctaagcgtt	ttggtagctg	tgtttccctg	catatagctt	gcctcaaaaag	aaaatttttaa	840
gtttgatttt	gtattaaatt	taaacaaaata	raggcgggtt	gcagtggctc	acgcctgtag	900
tcccagcact	ttggtaggcc	aagggtggcg	gatcacctga	ggtcaggatt	tcgagaccag	960
cctggccaat	atgggtgaac	cccattctta	ctaacatata	aaaatttagct	gggcatgggtg	1020
gcacatgcct	gtaatcctgg	ctacttggga	agctgaggca	ggagaatctc	ttgaaccag	1080
gaggcggagg	ttgcagttag	ccgagattgt	gccactgcac	tacaacctgg	acaacagagc	1140
aagattccgt	ctcaaaaaaa	aaaaaaaaaa	ctcga			1175

<210> 105

<211> 1021
 <212> DNA
 <213> Homo sapiens

<400> 106
 ggcacgagaa taaaatagat ttaggggtatt cgacac'at'tt aatacataat gaatagtga 60
 ttctctgaat aatataatgc ctaaaagtga ctgatatgca aaatatttaa atcccttaca 120
 ggatgtgcat ttaccaatca gaacagatgt tggctttact gcttgtatta gttttctgta 180
 ttagtttgct tgtattagtt tgctggggca gccataacaa agtaccacag aaatttatat 240
 tctcacaatt ctgggggcta gaagactgag ataaagatgt ctgcagggtt ggtttcttct 300
 aaggactctc ttgttggtt gtagatgggt gttgtctccc tgtgtcttca cctgcacttc 360
 cctctgtacc tgcctgtatc ttaatctctt cttataagga cagagtcata ttggattagg 420
 gcttcacctt aatgacctcc ttttaacctt attacctctt tagagactct gcttcaaat 480
 acaggcacat tctgaggtac ttttaagggtt gaacttcaac atatgcattt ggaactgggg 540
 cataactgag tccatgacac tattggagag ggttctggag ccataacaga tggaaatgaat 600
 gggccttgct tggattctaa tttaaataaa ccaactatag aaaaacatgt tttaggcaat 660
 cagat'tttt' gttatgggct aaatatttag aagatgttaa gaaatgttta ttaatttcat 720
 taggtaaaat aatgttggtg tggttatgta gagtatgact tacgtgcaca gaacagaatg 780
 ataagggtgc caggatttgc tttaaaatat tctagat'ttt' tttaaaagg' gagaggaaat 840
 gacacatgta tatctaaatt ttgatgggtg ttgaagctgg ctggtaggta cctagtattt 900
 cattatactt tttcctttgt gtgttggaa attttataa taagcaggaa aaataaaagt 960
 aagaatatat gtattataga taatttgaat tatgtagaat catgaaaaaa aaaaaaaaaa 1020
 a 1021

<210> 107
 <211> 830
 <212> DNA
 <213> Homo sapiens

<400> 107
 ggcacgagta aggactgtgt tttttatgca tttcttgatc caggcatggc agttcctctt 60
 ttctctgtaca tattcacact cctgccactt ctaccttctt tcttatccct ctgcttttca 120
 cctctgactg taaaaagaag tagcagttcc gaaagcaaga gtccctatg aacacggaag 180
 aagacattgg caacttttga gtacaacaac tataattaat agagtaattt aagaacatca 240
 gccagtgaat tttatacaag atagtgaag agaaaaggaa gattaaattag gggtagt'tta 300
 ggatgccatt aaatagccta gaattagggg agtagtcgtt gaatagaaaag gaggccacaa 360
 atttgaggga tataagctaa gaattggtaa gccaaagaag agcaaaaagg ttgggcagta 420
 aggataatga ggaacaaaat agagaactca gaagcaatat ctgactgtta tcattggaag 480
 aatttttttg cttgcttgag gctggatatt gaagtgatc aggatacttg agtgactatc 540
 tgatgggctt ttggaactag ctctcaagag gtgaaaatta gctttttttt cttttctttt 600
 cttttttttt ttttttgagg caaggctcca ctgttggtga ggctgaacct cctgggctca 660
 agcagttgtc ccattgcagc ctctcagat actctgtaag ccaaggcagg gggaaatatt 720
 tgtgtcag' agtttgaggc tgtggtgagc taagatcaca ctgctgtgct cacttcagcc 780
 tgggcaacac agtgaaacct cgtctccatc tgtttaaaaa aaaaaaaaaa 830

<210> 108
 <211> 1301
 <212> DNA
 <213> Homo sapiens

<400> 108
 aagaggggtg ctgccaggtt cttgtctgtc ctgtgctctg agagtgggtga gttatggaga 60
 cccaggtccc agcccacccc acctcagcca ggcttgaca ctgacctctg ccttgccccct 120
 cagtgccccg attcatcaag tacacaggct atgggaatgc tgctggcctt ctggctgcca 180
 ggggcctcat ggcaggaggc cggccccagg gccagtactc agaggatgag gacacagaca 240
 cagatgagta caaggaagcc aaagccagca taaacctgt gaccgggagg gtggaggaga 300
 agccgcctaa ccttatggag ggcattgacag aggagcagaa ggagcacgag gccatgaagc 360

tggtgaccat	gtttgacaag	ctctccagga	acagagtcac	ccagccaatg	gggatgagtc	420
cccgggggtca	tcttacgtcc	ctgcaggatg	ccatgtgcga	gactatggag	cagcagctct	480
cctcggaccc	tgactcggac	cctgactgag	gatggcagct	cttctgctcc	cccatcagga	540
ctggtgctgc	ttccagagac	ttccttgggg	ttgcaacctg	gggaagccac	atccccactgg	600
atccacaccc	gccccactt	ctccatctta	gaaacccctt	ctcttgactc	ccgttctgtt	660
catgatttgc	gtcttggtcca	gtttctcatc	tctggactgc	aacgggtcttc	ttgtgctaga	720
actcaggctc	agcctcgaat	tccacagacg	aagtactttc	ttttgtctgc	gccaagagga	780
atgtgttcag	aagctgctgc	ctgagggcag	ggcctacctg	ggcacacaga	agagcatatg	840
ggagggcagg	ggtttgggtg	tgggtgcaca	caaagcaagc	accatctggg	attggcacac	900
tggcagagcc	agtgtgttgg	ggtatgtgct	gcacttccca	gggagaaaaac	ctgtcagaac	960
tttccatacg	agtatatcag	aacacaccct	tccaaggtat	gtatgctctg	ttgttcctgt	1020
cctgtcttca	ctgagcgcag	ggctggagge	ctcttagaca	ttctccttgg	tcctcgttca	1080
gctgcccact	gtagtatcca	cagtgcccca	gttctcgctg	gttttggcaa	ttaaaacctcc	1140
ttcctactgg	tttagactac	acttacaaca	aggaaaatgc	ccctcggtgtg	accatagatt	1200
gagatttata	ccacatacca	cacatagcca	cagaaacatc	atcttgaaat	aaagaagagt	1260
tttgacaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaactcg	a		1301

<210> 109

<211> 1932

<212> DNA

<213> Homo sapiens

<400> 109

aatttttttt	tttttttttt	tttttttgaa	aaaaaaaaatg	ggtagtgtat	attttgcagg	60
tttaagacaa	ctcaggacaa	taaaaacaat	ggactttaca	tgtgtatata	tatagctctc	120
ttaggcacca	taatcagtat	gagccaacaa	tatttaaaact	tgattcaggc	cacattcaga	180
catttgcctc	tatatacaaa	tatttaaaatt	aaatacaaat	tgaaatgtgt	tctgtttacat	240
acaaaaaagg	aaaaactata	caacgcagag	cagtgtgtgt	gttttaaaata	attacattta	300
catgtaagct	aaatggaacc	agcaatgggtg	ctcaagtttt	tatcatccct	tccagaaaat	360
ctttttctac	catctcttct	attttttgcc	tggctttgct	ggaacatggg	ttgtggttct	420
ccagtttcat	gtccttatta	gggaaggcat	ttgagtagag	gataggactc	cctgagtgtc	480
ctccacatcg	gcttgtgact	ttgctgttga	agacttgact	gagcacattg	aagaacggca	540
ggagctgctc	catactgcgc	acgggtgcaga	tggtagcag	caagtgcctt	ggctcccaac	600
ccaatgttct	ccctgagttg	tcttctctctg	gatttttctg	cagaaaaaca	aaagtgaact	660
ggtattaata	caacagacaa	tgtggtatgt	tagaaaaat	aaaaatata	aaactttggc	720
aattgggtcaa	gaaatgaata	caaatgacat	taagttttta	actcctgacc	tgatcaaaac	780
ccttgggtgct	tctgagacct	tttactgcca	tttatagtt	ttacatggag	cagtctaaca	840
ttgtagtaat	agttcccaac	tagaatgcgc	agataagctt	agttaacaga	aatagctttg	900
aacaggaata	gagtcaaaaca	taaaagtctt	atgttgtgct	ttgtatttac	tcaaaaagct	960
cccagggtttc	tgaaccctca	ctactgtaac	caaggactag	gtcacaaaat	tactacagaa	1020
aaaaggaaca	aagtgtctta	tacatttcat	aatatatccc	cttttattat	aattagttaa	1080
ttccctttta	tctaaatggc	ctaaatttgc	catgatggta	gcagtgtcca	aagtgaataa	1140
ttactgtcag	tactgcatca	cagagaaaagg	aagggatccc	tcaggagaca	ctgctgtctc	1200
cttctgggtt	gtgctaaaca	acatagggag	gaaagctgga	cctggagtca	aaggaattga	1260
gttagtgtgc	tggctctgct	atacttacgg	caccttggg	caggatatac	aaaggttcct	1320
cacttataaaa	atgggacagt	ctaaaactac	cttttagtag	agaagtcaaa	tgagaaggta	1380
tgtgaaaact	ctgtcaacta	aatataaaga	ctaataattt	gggtatttaag	aggctagttt	1440
gagaagccac	ctgaattaca	caaacacagc	tacagacatc	attctgtcta	gagaaaagata	1500
agagagaaca	ggttgggtga	acttgggcag	aatcacagat	acaattccac	actaaagaat	1560
gaaaataaag	aatgaactag	acagaaggaa	gaaatcataa	agacttagga	agcagaatta	1620
caatctgtca	tatttaacaaa	tggagtcttc	cttctaagat	cagatgttgc	tcagaaactt	1680
tcattgttta	cctaataatt	taatatcact	agtttccctag	tgggtcaagc	agatgcaaaa	1740
tccagcttat	tttcttctat	gtgctctcaa	gcttattgct	tatttttaaag	taaaatcctg	1800
aaaaaggaaa	atattaggtt	ggtgcaaacg	taattgcggt	ttttgcattg	ttgaaatttg	1860
ccgttttata	ttggagtaca	ttcttaaaaa	aatgtgggtta	tgttatataca	aaaaaaaaaa	1920
aaaaaactcg	ag					1932

<210> 110
 <211> 1534
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (1212)
 <223> n equals a,t,g, or c

<400> 110
 tcgaagatag gttcagagacg gtggatgttg cagctgatca tgcagttggg ttcggtgctg 60
 ctcacacgct gcccccttttg gggctgcttc agccagctca tgctgtacgc tgagagggct 120
 gaggcacgcc ggaagcccgga catcccagtg ccttacctgt atttcgacat gggggcagcc 180
 gtgctgtgcg ctagtttcat gtcctttggc gtgaagcggc gctgggttcgc gctggggggc 240
 gcaactccaat tggccattag cacctacgcc gcctacatcg ggggctacgt ccactacggg 300
 gactggctga aggtccgtat gtactcgcgc acagttgccca tcatcggcgg ctttcttctg 360
 ttggccagcg gtgctgggga gctgtaccgc cggaaacctc gcagccgctc cctgcagtc 420
 accggcccagg tgttcttggg tatctacctc atctgtgtgg cctactcact gcagcacagc 480
 aaggaggacc ggctggcgta tctgaaccat ctcccaggag gggagctgat gatccagctg 540
 ttcttcgtgc tgtatggcat cctggcccc cggcctttctg tcaggctact acgtgacct 600
 cgctgcccag atcctggctg tactgctgcc ccctgtcatg ctgctcattg atggcaatgt 660
 tgcttactgg cacaacacgc ggcgtgttga gttctggaac cagatgaagc tcttgggaga 720
 gagtgtgggc atcttcggaa ctgctgtcat cctggccact gatggctgag ttttatggca 780
 agaggctgag atgggcacag ggagccactg agggtcaccc tgccttcctc cttgctggcc 840
 cagctgctgt ttatttatgc tttctggctc gtttgtttga tcttttgctt ttttaaaatt 900
 gttttttgca gtttaaggag agctcatttg tccaaatttc tgggcttcag cgcttgggag 960
 ggcaggaacc ctggcactaa tgctgtacaa ggtttttttc ctgttaggaa gaacttgagg 1020
 ccagctgccc actgagtctt ctgtccctga agaaaggag tatgggcag ggcctgggat 1080
 ccggctactg agagtgggag agtgggagac agaggaagga agatggagat tgggaagtga 1140
 caaatgtgaa aaattcctct ttgaacctgg cagatgcagc taaactctgc agtagtgttt 1200
 ggagactgtg anagggagtg tgtgtgttga cacatgtgga tcaggcccag gaagggcaca 1260
 ggggctgagc actacagaag tcacatgggt tctcagggtg tgccaggggc agaaacagta 1320
 ccggctctct gtcactcacc ttgagagtag agcagacctt gttctgctct gggctgtgaa 1380
 ggggtggagc aggcagtggc cagctttgcc cttctgtctg tctctgttct tagctccatg 1440
 gttggcctgg tgggggtgga gttccctccc aaacaccaga ccacacagtc ctccaaaaat 1500
 aaacatttta tatagacaaa aaaaaaaaaa aaaa 1534

<210> 111
 <211> 2871
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (1234)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1259)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1283)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1284)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1287)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1378)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1912)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1913)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1935)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (1947)
 <223> n equals a,t,g, or c

<400> 111
 ggggcgaggc ggcgggtggtg gctgagtcgg tgggtggcaga gggaaaggcg acagctctag 60
 ggggttggcac cggccccgag aggaggatgc ggggtccggat agggctgacg ctgctgctgt 120
 gtgcggtgct gctgagcttg gcctcggcgt cctcggatga agaaggcagc caggatgaat 180
 ccttaggatt ccaagactac ttgacatca gatgagtcag taaaggacca tactactgca 240
 ggcagagtag ttgctggtca aataatttctt gattcagaag aatctgaatt agaattcctct 300
 attcaagaag aggaagacag cctcaagagc caagaggggg aaagtgtcac agaagatatc 360
 agctttctag agtctccaaa tccagaaaac aaggactatg aagagccaaa gaaagtacgg 420
 aaaccaggtg gtctggacat ttcccttgct ttttgattta tttaggggac aactgaaaat 480
 tttaaagctaa tgaataaaga agctgaagaa gactggcttc actgattatt acccacaat 540
 aatatatgga gtgtagtttg gagagaattt ctgattttta atataccaaa gttatttact 600
 taacagattt gacccaagtt acataagctg aattccatag atgagattgc atttatcttt 660
 ttattaatat cctgacgttt ctttttggaa tcaagtatgt aaataccttt ctacttgaga 720
 aatttaaaaa aataatatgt cagggtagca tcaaaaacag aaactcagga agaactgtct 780
 aacatgccac tatggttatt ttcaggattg tggtaggatt attactacaa taaccttgct 840
 gatcatccta ggtaatttga tttaggtgtt tatccctcca gcttgacgtt tgtgtgtact 900
 tccatcctat tctaactttc acaattttga aaactctact tttcaattta aaacttgacc 960
 tttttcacaa tagtcttcta aataattaatc ctttttaaaa ttattaatgt atgcattcta 1020
 ccatgtcttg ttgcccttct gtcatttatt atttcaatga tcaactgcctt atctgtttat 1080
 tattccatta ctaaccactc tcaggaactg atccaaattg gaacttttta aaaaaataga 1140
 attttttttt actacaaata taccattctc atgaatagaa aactattttt aaaaaattaa 1200
 aagtacctga atcccatcaa ctagaaataa cagntgctca catcatggag aatattctnt 1260
 caggattttt ttaggtgtat cannagnttt aagaaaagta aattgggtca aaatgtgtat 1320
 tctatgttgc agcctgcatt ttgcacttaa cagtttacca tgaatgtttt tccatgtnat 1380

ttagcttatt	ttcaatttga	tagttaatgg	ctataaaaaa	ttttatttgt	aaatatatgg	1440
taccataaac	caaaacgttt	atgttttgc	gggagatcat	tttagatgta	tctttgtgcc	1500
tatctgtgaa	aatttcctta	gaattcttaa	agtaaatttg	ctgaattgga	tgtaaaagtt	1560
taaaagacag	tagatacaca	ttgttaaatt	gttcttcaga	aaagtgattc	tgtttacttc	1620
ctctgaaaat	acctctttcc	ctgtacctca	cccaagtgtg	tattaccatt	ttaaaatttt	1680
tactaaatat	aatttaattc	attcttcat	ttttaattag	ggattttttt	tttttttttt	1740
tagtttttat	atttttagag	acatgggtct	actctgtcac	tcaggctgga	gtgcagtggc	1800
acgatcatag	ttcactgcag	ccctgaactc	agggtgatct	cctctttggc	ctcccaaagt	1860
gctgggatta	caggcatgag	acactacatc	aagcctggga	aatttttaaa	cnngcataaa	1920
aagtagaggg	taaantgcc	ccccttncca	accctcactc	agcttcaata	aacatcaaca	1980
ttctgctggt	cttttcatct	ttactacct	acacacattt	gccctttttt	tctttcctag	2040
aatattttaa	ggcaaattcc	gcatgtgctt	ttacatctct	gtgtctctct	ctagtaagaa	2100
tgtctttttc	agcaaaatca	cagaattagc	cattactcct	ttattacct	ataccagac	2160
tgttttaatt	ttctagtgtc	ttcaagtgtc	tgtttacaat	ttatttgttt	gaattcattt	2220
ctaagctact	caggaggctg	agtcaggaga	attgcttgaa	ccctagaggc	agagggttgc	2280
gtgagctgag	atcgtgccac	tgtactccag	cctgggtgac	aaagtgagac	tccatgtcat	2340
aacaaaacaa	aacaaaacaa	aacatgacga	aacaaaacaa	acaattgaat	aatatccctc	2400
agttattatt	tagagaagag	atgactcaat	ataaatgtaa	tatgtgatcc	atttgggtcac	2460
agaatgaatt	tcagtcattg	cttaggtaag	cctggtaaaa	aaaaaaaaaa	cgggattcca	2520
tgattagtgt	gttgaaaaag	gggtttgccc	aaagtgtaac	agacagagcc	gggtgcagtg	2580
gctcacgcct	gtaatcccg	aactttggga	ggccaaggca	ggcggatcac	gagggtcaaga	2640
gattgagaca	atcctggcca	acatgggtga	gccctgtatc	aactaaaaat	acaaaaacaa	2700
ttagctgggc	gtgggtgacat	gtgcctgtay	lcccgactac	tcaggaggct	gaggcaggag	2760
aattgcttga	acctgggagg	cggagggtgc	agtgagctga	gattgtctca	ctgcactcca	2820
gcctggccac	agagtggagat	tccatctcaa	aaaaaaaaaa	aaaaaaaaaa	a	2871

<210> 112

<211> 1037

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (936)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (946)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (951)

<223> n equals a,t,g, or c

<400> 112

ccggtccgga	attccccggg	cgaccacgc	gtccggggcg	gcgcaggacg	tgcactatgg	60
ctcggggctc	gctgcgcggg	ttgctgcggc	tcctcgtgct	ggggctctgg	ctggcggttc	120
tgcgctccgt	ggccggggag	caagcgccag	gcaccgccc	ctgctcccgc	ggcagctcct	180
ggagcgcgga	cctggacaag	tgcattggact	gcagcacctc	ctgccccctt	ccggctgctt	240
tggccccatc	ttggggggcg	tctgagcctg	accttcgtgc	tggggctgct	ttctggcttt	300
ttggctctga	gacgatgccg	caggagagag	aagttcacca	cccccataga	ggagaccggc	360
ggagagggct	gcccagctgt	ggcgctgac	cagtgcacat	gtgccccctg	ccagccgggg	420
ctcgccact	catcattcat	tcatccattc	tagagccagt	ctctgcctcc	cagacgcggc	480
gggagccaag	ctcctccaac	cacaaggggg	gtggggggcg	gtgaatcacc	tctgaggcct	540
gggcccaggg	ttcaggggaa	ccttccaagg	tgtctggttg	ccctgcctct	ggctccagaa	600
cagaaaaggga	gcctcacgct	ggctcacaca	aaacagctga	cactgactaa	ggaactgcag	660

catttgcaca	ggggaagggg	gtgccctcct	tcttagaggg	cctggggggc	aggtgactt	720
ggggggcaga	cttgacacta	ggccccactc	actcagatgt	cctgaaattc	caccacgggg	780
gtcaccctgg	gggggttagg	acctattttt	aacactaggg	ggctggccca	ctaggagggc	840
tggccctaag	atacagaccc	ccccaaactcc	ccaaagcggg	ggggggatat	ttattttggg	900
gagagtttgg	aggggagggg	gatttttttt	ttaaangatt	ttttantttt	naaaaaaaaa	960
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1020
aaaaaaaaaa	aaaaaaa					1037

<210> 113

<211> 2214

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (289)

<223> n equals a,t,g, or c

<400> 113

cggacgcgtg	ggtcgaccca	cgcgtccggg	aaaaarggaa	aaaratgccgt	gtaaaatctc	60
gttctgtgtc	tgaattgccg	taggctcaga	tcttcatttg	aggttctgtg	tctgaattgc	120
cgtaggctca	gatcttcatt	tgaggttatg	ttctataagt	taacgctgat	cttgtgtgag	180
ctttcggtag	ctggagtaac	acaggcggcc	tcacagcgac	ctctccagcg	ccttccaagg	240
cacatctgca	gccagcgtam	tcctcctggg	agatgcctcc	tcaaggccnt	gctccagacc	300
acgtgggrar	ggcctgacaa	gccaattccc	aggctgtccc	cacccttgra	gagtgacct	360
aaacgctaga	cagatgggga	atgggaaaga	aaagaaaagc	gcagacctca	agttaaaatt	420
ccctcaaaaa	cgttttttatt	tatctgcttt	ttctgaaagg	ataaaggctt	tttgaatttt	480
attttctaac	aaataacatg	aacacttcta	gaaacccatg	aaaaacacaa	agtattcaaa	540
atagaaagaa	aaattaccca	ttactcttta	agccagcatt	atccattgcg	gtgcttttgg	600
agttgggtga	ggccgtagcc	tctgccaagt	caaggagccc	gggtggtggc	gtggcattcc	660
tgcagggttg	tttttttttc	tttgagatgg	agtctcactc	ttgtcacccc	agctgggaatg	720
tgggtggtga	aacagctcac	tgcagccttg	accctgaggc	tcaagcgatc	cttctgcctt	780
ggcctcctga	gtagctggga	ttccaggcga	gagtcaccac	accctgtcca	tgttcctgca	840
ggctcttgata	tgcgaggacg	ctgtgtcttc	cctgccacat	ttcttcttcc	tttcttgaga	900
cagacccttg	ctccatcacc	caggccagag	tgtggtsgtg	cgaacacggc	tcactgcagc	960
ctcgaccctc	aggctcaagc	gatccctcacg	cctcggaacc	ccaaagtgtc	gggatcacag	1020
gcgagatgca	ccatgtctggc	ctgaattctc	agggtatttr	cggttgargt	gycacttact	1080
tarccatccc	gttttcaaga	gtgtagggtg	tcaccctgtc	tctgccgtcg	acctggccctg	1140
gaccctcggc	tgtgagaggg	aggggtgggc	tgggctggag	gaacctraag	ccctcgtgat	1200
gtcacaagcc	catctggctg	ggcatccccc	gctgtgtctc	gagctgcaca	tgccccaggc	1260
ggccccccaca	gcagaggcga	gccactgrag	ggtgragggc	ttccacggac	ggtcttcagg	1320
ggragaagaa	gggcccaggc	ccccaggaga	ctcaggagac	cagagcctgg	ggtcaggggc	1380
tmagcagggg	ctyarccagg	gctggatgtc	cggagccagc	cccgmagccc	tgkgktcttt	1440
gttcttccga	ctcccaccgt	cogtgtgaac	agctccagcc	ccacctgcgc	ctccctgtgc	1500
tgggctccat	caggagagccc	agaagacgtg	tgtgtctctg	aaattgggtc	cctacatgcc	1560
tttgtcccag	tgcaccttgc	tccttccatt	tactatcgag	atttaaatgc	ctgttttctc	1620
cccagagggt	gacggatata	ttcagacgtt	acgacacgga	tcaggacggc	tggtattcagg	1680
tgtcgtacga	acagtacctg	tccatggtct	tcagtatcgt	atgaccctgg	cctctcgtga	1740
agagcagcac	aacatggaaa	gagccaaaat	gtcacagtcc	ctatctgtga	gggaatggag	1800
cacaggtgca	gttagatgct	gttcttccct	tagattttgt	cacgtgggga	cccagctgta	1860
catatgtgga	taagctgatt	aatggttttg	caactgtaat	agtagctgta	tcgttctaatt	1920
gcagacattg	gatttgggtga	ctgtctcatt	gtgccattag	gtaaatgtaa	tgtttcaggc	1980
attctgcttg	caaaaaaatt	tatcatgtgc	ttttctagat	gtctctgggt	ctatagtgca	2040
aatgctttta	ttagccaata	ggaattttta	aataacatgg	aacttacaca	aaaggctttt	2100
catgtgcctt	acttttttaa	aaaggagttt	attgtattca	ttggaatatg	tgacgtaagc	2160
aataaaggga	atgttagacg	tgtaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaag	2214

<210> 114

<211> 3300

<212> DNA

<213> Homo sapiens

<400> 114

tcgaccacg	cgccgggtga	gaggcaagga	cttttcacat	ttagggcatt	ctgaccacgt	60
cctgcttcag	agagattgtt	cccggcgtct	cagtgcctatg	gggagcaggt	tcctcctggt	120
cctgctctca	ggctctactg	tcttactggc	tctgccagga	tcagaagcca	agaattctgg	180
agcttcctgt	cctccatgcc	ctaaatatgc	cagctgccac	aacagcaccc	actgtacttg	240
tgaagatggc	tttcgggcca	ggctctggcag	gacatacttt	catgattcct	ctgagaagtg	300
tgaagatatt	aatgaatgtg	aaaccgggct	ggcaaaagtgc	aaagtataaa	catattgtag	360
gaataaaagt	ggagggttaca	tctgtagctg	tttggtaaaa	tatactttat	tcaactttct	420
ggctgggtatt	atagattatg	atcatccgga	ttgttatgag	aacaatagtc	aagggacgac	480
acagtcaaac	gtggatattt	gggtgagtg	gggaagcct	ggatttgga	aacagctggt	540
acgtataact	atgccatttt	cctaccctaa	cattaacatg	tcttcctgtg	atcttttaggg	600
tagggtagtt	ctatccaggg	gtaattttgt	cctctgtccc	aaggctcatct	gtcaatgact	660
ggggacactt	ttggttgtca	taccttgggg	gtgatgtgtg	tgactggcat	ctgggtggatg	720
gagaccaggg	atacagctca	acatcctaca	gtgccagga	cagcctccca	caatcaagaa	780
gtgcttagtg	ccatatgtcc	atagagatag	agaaatacaa	gtgtaggggg	aaagtgcctc	840
agctggcatt	caaagacctt	catcaagcac	ctagattctc	aatgccacac	gcaccttgta	900
gcaccttaat	aaatatcggt	gccttctggg	ctccctactc	caacacttgt	gcattatccc	960
tatttctttt	atttcagtaa	gagtcacact	aatttagctta	atttttttgg	aaggaaaaatc	1020
tgagaagaaa	tgggaagcaga	gaggactttg	caagaagggg	tactcaacta	attcaaaagcg	1080
tggagttgag	catctgggaat	gcgagttttg	cttctccagg	aaagggtcaa	atttctgaaat	1140
ttgatatagt	ctatgaaacc	aagaggtgca	atgagacaag	ggagaatgct	tttctgggaag	1200
ctggaaataa	caccatggat	atcaactgtg	ctgagtcttt	aaaaggaaac	ctaagagaga	1260
gcactgcagt	tgccttatca	cttatcaatc	tcttggggat	attctgaaatg	catccttttt	1320
tagtaaacga	aaagggatgc	aggaagtaaa	actgaactct	tacgtttgtga	gcggcaccat	1380
cgggttgaa	gaaaaaattt	ccctctctga	acctgtgttc	ctgacttttc	gccataatca	1440
gcctgggtgac	aagagaacaa	aacatatctg	tgctacttgg	gagggatcag	agggaggccg	1500
ctgggtccacg	gagggctgct	ctcatgtgca	cagcaacggg	tcttacacca	aatgcaagtg	1560
cttccatctg	tccagctttg	ccgtcctcgt	ggctcttggc	cccaaggagg	acctgtgct	1620
gaccgtgatc	acccaggttg	ggctgacct	ctctctgctg	tgctcttctc	tggccatcct	1680
caccttcctc	ctgtgccggc	ccatccagaa	caccagcacc	tccctccatc	tagagctctc	1740
cctctgcctc	ttcctggccc	acctcctgtt	cctgacgggc	atcaacagaa	ctgagcctga	1800
ggtgctgtgc	tccatcattg	cagggtctgt	gcacttcttc	tacctggctt	gcttcacctg	1860
gatgctcctg	gaagggctgc	acctcttctc	caccgtcagg	aacctcaagg	tggccaacta	1920
caccagcacg	ggagagattca	agaagaggtt	catgtacct	gtaggctacg	ggatccacgc	1980
tgtgattatt	gctgtgtcag	caatagtttg	acccagaaat	tatggaacat	ttactcactg	2040
ttggctcaag	cttgataaag	gattcatctg	gagcttcatg	gggccagtag	cagtcattat	2100
cttgataaac	ctgggtgtct	acttccaaat	tctgtggatt	ttgagaagca	aactttcctc	2160
cctcaataaa	gaagtttcca	ccattcagga	caccagagtc	atgacattta	aagccatttc	2220
tcagctattt	atcctgggct	gttcttgggg	ccttgggttt	tttatgggtg	aagaagtagg	2280
gaagacgatt	ggatcaatca	ttgcatactc	attcaccatc	atcaacaccc	ttcagggaat	2340
gttgctcttt	gtggtaacct	gtctccttaa	tcgccagggt	cgaatggaat	ataaaaagtg	2400
gtttagtggg	atgcggaag	gggtagaaac	tgaaagcact	gagatgtctc	gctctactac	2460
ccaaaccaaa	acggaagaag	tggggaagtc	ctcagaaatc	tttcataaag	gaggcactgc	2520
atcatcatct	gcagagtcaa	ccaagcaacc	gcagccacag	gttcatctcg	tctctgctgc	2580
ttggctaaag	atgaactgac	ctggcaagtg	ccatggcaat	gaccgggaag	ttacccctcc	2640
tttccgtttg	tctacagcgc	ccctgtgggc	acacatagat	tggacaaatg	ccactatttc	2700
tagctttctc	gtgaaaagtc	taggtctatt	cacctatttt	ggctttttat	gttcatagaa	2760
agaacaagac	atttgggaga	attcttagat	ccagagtcca	gtagtgtggc	acgtgcaatg	2820
aagtgtcgga	aggatgcatt	ttaaagatgg	cgggcgggag	aagtggattt	tcttcttgca	2880
gctactgcca	ccttgccaga	aactcactaa	ctggcatctg	gattcagctc	atagtccct	2940
ttctggcctc	tctgctgtat	ttcatgtctc	caaagatctt	acattaacac	tccacattca	3000
cataattcaa	caattttcat	atggatcagt	attaaagagg	gtgttgcat	ttgcaataca	3060
aaaatgcatt	atcaggtgct	ggagaggatg	tggagaaata	ggaacacttt	tacactgttg	3120
gtgggactgt	aaactagttc	aacctcctgt	gaagtcagtg	tggcgatttc	tcagggatct	3180

agaactagaa ataccatttg acacagctat cccattactg ggtatatacc ccaagggcta 3240
 taaatcatgc tgctataaag acacatgcac acgtaaaaaa aaaaaaaaaag ggcggccgct 3300

<210> 115
 <211> 1286
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (1149)
 <223> n equals a,t,g, or c

<400> 115
 gactgcttta cggacatttg atgaagccga agcatttaga atgggtgcctg gcacacagtt 60
 ggtgcgtgat atgggtaagc tttgtgtccc caccacatc tcattctgaa tgtgacggtt 120
 .tccccggctc cctcctgccg ccattgtgaag aagggtcgtt cttccccctc accttccacc 180
 accatgattg ccattggatgc tccccactcc aaagcagccc tggacagcat taacgagctg 240
 cccgagaaca tctgtctgga gctgttcacg cactgtcccg cccgccagct gctgctgaac 300
 tgccgccttg tctgcagcct ctggcgggga cctcatcgac ctcatgacct tctggaaacg 360
 caatgcctgc gagagggcct catcaccaag gactgggacc agcccgtggc cgactggaaa 420
 atcttctatt tcttacggag cctgcatagg aactcctgcg caaccctgtg gctgaagagg 480
 atatgtttgc atggcaaatt gatttcaatg gtggggaccg ctggaagggt gagagcctcc 540
 ctggagccca cgggacagat tttcctgacc ccaaagtcaa gaagtatttt gtcacatcct 600
 acgaaatgtg cctcaagtcc cagctggtgg accttgtagc cgagggtctac tgggaggagc 660
 tactagacac attccggccg gacatcgtgg ttaaggactg gtttgctgcc agagccgact 720
 gtggctgcac ctaccaactc aaagtgcagc tggcctcgac tgactacttc gtgttggcct 780
 ccttcgagcc cccacctgtg accatccaac agtggaaaca tggcacatgg acagaggtct 840
 cctacacctt ctacagactc ccccggggtg tccgctacat cctcttccag catgggggca 900
 gggacaccca gtactgggca ggctggtatg ggcccagat caccaacagc agcattgtcg 960
 tcagccccaa gatgaccagg aaccaggcct cctccgagge tcagcctggg cagaagcatg 1020
 gacaggagga ggctgcccac tgcacctacc gagctgtgtt ccagattttc tgacagctgt 1080
 ccattcctgt tctgggtcag ccagaggttc ctccaggcag gagctgagca tgggggtggc 1140
 aagtgaggnc cctgtaccaaa gcgacttctg ccccggttca acctaccaaa gcttggggga 1200
 acttactgca catagctctg acgttttgtt gtaataaatg ttttcaggcc gggcaaaaaa 1260
 aaaaaaaaaa aaaaaaaaaa aaaaaa 1286

<210> 116
 <211> 2189
 <212> DNA
 <213> Homo sapiens

<400> 116
 ggaattcggc acgaggcgcc cgaggatgtg ctgctggccg ctgctcctgc tgtgggggct 60
 gctccccggg acggcgggcg ggggctcggg ccgaacctat ccgcacggga cctccttgga 120
 ctcgaggggc aagtactggc tgggctggag ccagcggggc agccagatcg ccttccgccc 180
 ccagggtgcg actgcaggct acgtgggctt cggtctctcg cccacggggg ccattggcgtc 240
 cgccgacatc gtctgtggcg ggttggccca cggcgggccc tacctccagg attattttac 300
 aaatgcaaat agagagttga aaaaagatgc tcagcaagat taccatctag aatatgccat 360
 ggaaaatagc acacacacaa taattgaatt taccagagag ctgcatacat gtgacataaa 420
 tgacaagagt ataacggata gcactgtgag agtgatctgg gcttaccacc atgaagatgc 480
 aggagaagct ggtcccaagt accatgactc caataggggc accaagagtt tgcggttatt 540
 gaatcctgag aaaactagt tgcctatctac agccttacca tactttgatc tggtaaatca 600
 ggacgtcccc atcccaaaaca aagatacaac atattggtgc caaatgttta agattcctgt 660
 gttccaagaa aagcatcatg taataaaggc tgagccagtg atacagagag gccatgagag 720
 tctggtgcac cacatcctgc tctatcagtg cagcaacaac tttaacgaca gcgttccctg 780
 aatccgggca cgaattgcta tcccccaac atgcccagtg cattcctcac ctgtgaaact 840

gtgatttttg	cctgggctat	tggtggagag	ggcttttctt	atccacctca	tgttggatta	900
tcccttggca	ctccattaga	tccgcattat	gtgctcctag	aagtccatta	tgataatccc	960
acttatgagg	aaggcttaat	agataattct	ggactgaggt	tattttacac	aatggatata	1020
aggaaatatg	atgctggggg	gattgaggct	ggcctctggg	tgagcctctt	ccataccatc	1080
cctccaggga	tgcctgagtt	ccagtctgag	ggctactgca	ctttggagtg	cctggaagag	1140
gctctggaag	cgaaaaagcc	aagtgggaatt	catgtgtttg	ctgttcttct	ccatgctcac	1200
ctggctggca	gaggcatcag	gctgcgtcat	tttcgaaaag	ggaaggaaat	gaaattactt	1260
gcctatgatg	atgattttga	cttcaatttc	caggagtttc	agtatctaaa	ggaagaacaa	1320
acaatcttac	caggagataa	cctaattact	gagtgtcgct	acaacacgaa	agatagagct	1380
gagatgactt	ggggaggact	aagcaccagg	agtgaatgt	gtctctcata	ccttctttat	1440
taccaagaa	ttaatcttac	tcgatgtgca	agtattccag	acattatgga	acaacttcag	1500
ttcattgggg	ttaaggagat	ctacagacca	gtcacgacct	ggcctttcat	tatcaaaagt	1560
cccaagcaat	ataaaaacct	ttctttcatg	gatgctatga	ataagtttaa	atggactaaa	1620
aaggaaggtc	tctccttcaa	caagctggtc	ctcagcctgc	cagtgaatgt	gagatgttcc	1680
aagacagaca	atgctgagtg	gtcgattcca	aggaatgaca	gcattacctc	cagatataga	1740
aagaccctat	aaagcagaac	cctttgggtg	gtggcacgtc	ttcttctctt	tccctgcaca	1800
gagatttctc	ccatcaactt	gcttgtttgc	cttctgctac	tcagctgcac	gctgagcacc	1860
aagagcttgt	gatcaaaaatt	ctgttggact	tgacaatgtt	ttctatgatc	tgaacctgtc	1920
atttgaagta	cagggttaag	actgtgtcca	ctttgggcat	gaagagtgtg	gagacttttc	1980
tccccatttt	tccctccctc	cttttctctt	tccatgttac	atgagagaca	tcaatcaggt	2040
tctcttctct	ttcttagaaa	tatctgatgt	tatatataca	tggtcaataa	aataaaactg	2100
gcctgactta	agataaccat	tttaaaaaat	tgggctgtca	tgtgggaata	aaagaattct	2160
ttctttctta	aaaaaaaaaa	aaaaaaaaaa				2189

<210> 117

<211> 1763

<212> DNA

<213> Homo sapiens

<400> 117

ggtggcctag	agatgctgct	gccgcggttg	cagttgtctg	gcacgcctct	gcccggcagc	60
cgcctccacc	gccgtagcgc	ccgagtgtcg	gggggcgcac	ccgagtcggg	ccatgaggcc	120
gggaaccgcg	ctacaggccg	tgctgctggc	cgtgctgctg	gtggggctgc	gggcgcgcac	180
gggtcgccctg	ctgagtgggc	agccagtctg	ccggggaggg	acacagaggc	cttgttataa	240
agtcatcttac	ttccatgata	cttctcgaag	actgaacttt	gaggaagcca	aagaagcctg	300
caggagggga	tggaggccag	ctagttagca	tcgaattctg	aaatgaaca	gaaactgata	360
gaaaakttca	ttgaaaacct	cttgccatct	gatgggtgay	tcggawtgg	gtcaggagg	420
cgtgaggaga	aacaaagcaa	tagcacagct	gccaggacct	ttatgcttgg	actgatggca	480
gcataccaca	atttaggaac	tggtatgtgg	atgagccgtc	ctgcggcagc	gaggtctgcg	540
tggtcatgta	ccatcagcca	tcggcaccgc	ctggcatcgg	aggccctac	atgttccagt	600
ggaatgatga	ccggtgcaac	atgaagaaca	atttcatttg	caaataattt	gatgagaaac	660
cagcagttcc	ttctagagaa	gctgaagggt	aggaacacaga	gctgacaaca	cctgtacttc	720
cagaagaaac	acagggaagaa	gatgccaaaa	aaacatttaa	agaaagttaga	gaagctgcct	780
tgaatctggc	ctacatccta	atccccagca	ttccccctt	cctcctcctt	gkggtcacca	840
cagttgtatg	ttgggttttg	atctgtagaa	aaagaaaaacg	ggagcagcca	gacctagca	900
caaagaagca	acacaccatc	tggccctctc	ctcaccaggg	aaacagcccg	gacctagagg	960
tctacaatgt	catwagaaaa	caaagcgaag	ctgacttagc	tgagaccctg	ccagacctga	1020
agaaatatttc	attccgagtg	tggtcgggag	aagccactcc	cgatgacatg	tcttgtgact	1080
atgacaacat	ggctgtgaac	ccatcagaaa	gtgggtttgt	gactctgggt	agcgtggaga	1140
gtggatttgt	gaccaatgac	atttatgagt	tctccccaga	ccaaatgggg	aggagttaagg	1200
agtctggatg	ggtggaaaaat	gaaatatatg	gttatttagga	catataaaaa	actgaaactg	1260
acaacaatgg	aaaagaaatg	ataagcaaaa	tcctcttatt	ttctataagg	aaaatacaca	1320
gaaggtctat	gaacaagctt	agatcaggtc	ctgtggatga	gcatgtgggt	cccacgacct	1380
cctgtctggac	ccccacgttt	tggtctgtatc	ctttatccca	gccagtcac	cagctcgacc	1440
ttatgagaag	gtaccttgcc	caggtctggc	acatagtaga	gtctcaataa	atgtcacttg	1500
gttgggttga	tctaactttt	aagggacaga	gctttacctg	gcagtataaa	agatgggctg	1560
tggagcttgg	aaaaccacct	ctgttttctt	tgctctatac	agcagcacat	attatcatac	1620
agacagaaaa	tccagaatct	tttcaaagcc	cacatatggt	agcacaggtt	ggcctgtgca	1680

tcggcaattc tcatatctgt ttttttcaaa gaataaaatc aaataaagag caggaaaaaa 1740
 aaaaaaaaaa aaaaaaactc gag 1763

<210> 118
 <211> 1375
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> (18)
 <223> n equals a,t,g, or c

<400> 118
 ggcggggcgc gcggaagngg cggckgcgcg gccggggcag ccattgtcgcc attgtctgcg 60
 gcgcggggcg ccctgcgggt ctacgcggta ggccgcgcgc tgatcctggc gcagctgctg 120
 cggcgctgcc gsgggggcct cctggagcca gtttcccccc cagcactga ccgtgtcgt 180
 atagtacgg gagggacaga tggcattggc tattctacag cgaacatctg gcgagacttg 240
 gcatgcattgt tatcatagct ggaaataatg acagcaaagc caaacaagt gtaagcaaaa 300
 taaaagaaga aaccttgaac gacaaagtgg aatttttata ctgtgacttg gcttccatga 360
 cttccatccg gcagtttgtg cagaagttca agatgaagaa gattcctctc catgcctga 420
 tcaacaatgc tgggggtgat atggtccttc agaggaaaac cagagatgga ttcgaagaac 480
 atttcggcct gaactaccta gggcacttcc tgctgaccaa ccttctcttg gatacgtga 540
 aagagtctgg gtccccctggc cacagtgcga ggggtggcac cgtctcctct gccacccatt 600
 acgtcgctga gctgaacatg gatgaccttc agagcagtgc ctgctactca cccacgcag 660
 cctacgccca gagcaagctg gcccttgtcc tgttcacctt ccacctccag cggctgttg 720
 cggctgaggg aagccacgtg accgccaacg ttgtggacct cggggtggtc aacacggacs 780
 tctacaagca cgtgttcttg gccacccgtc tggcgaagaa gcttctcggc tggttgcttt 840
 tcaagacccc cgatgaagga gcgtggactt ccatctacgc agcagtcacc ccagagctgg 900
 aaggagtgtg tggccgttac ctatacaacg agaaagagac caagtccctc cacgtcacct 960
 acaaccagaa actgcagcag cagctgtggg ctaagagttg tgagatgact ggggtccttg 1020
 atgtgacctt gtgatctcct gtctcaggat agctgctgcc ccaagaaaca cattgcacct 1080
 gccaatagct tgtgggtctg tgaagactgc ggtgtttgag tttctcacac ccacctgcc 1140
 acagggctct gtccctctagt tttgagacag ctgcctcaac ctctgcagaa ctcaagaag 1200
 ccaataaac attttggagg ataatacccc caagtgtgtc tcaaccataa actttgtgat 1260
 tccaaagtgc ccagtgttca cagggtccat aaataattac attttccaac ataaaaaaaa 1320
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaagggcg gccgc 1375

<210> 119
 <211> 1022
 <212> DNA
 <213> Homo sapiens

<220>
 <221> SITE
 <222> (937)
 <223> n equals a,t,g, or c

<220>
 <221> SITE
 <222> (990)
 <223> n equals a,t,g, or c

<400> 119
 ggcacgagag agccrgcggc ccgagccggg gccatggcga actgctgagc tgcgtcctag 60
 gcccccggct ctacaaaatc taccgggaga gggactctga aagggccccc gccagcgtcc 120
 ctgagacgcc aacggcagtc actgcccccc attccagctc ctgggatacy tactatcagc 180

cccggtgccct	ggagaaacat	gctgacagca	tectggcact	ggcttcagta	ttctgggtcca	240
tctcttatta	ctcctctccc	ttcgcttct	tctacttgta	caggaaagg	tacttgagtt	300
tgtccaaagt	ggtgccgttt	tctcactatg	ctgggacatt	gctgctactt	ctggcaggtg	360
tggcctgctc	cgaggcattg	gccgtggaac	caacccccag	taccggcagt	tcaccacat	420
cttggaagca	acacatcgga	accagtcttc	agaaaaaag	aggcagcttg	ccaactacaa	480
ctttgacttc	cggagctggc	cagtcgactt	ccactgggaa	gaaccagca	gccggaagga	540
gtctcgaggg	ggcccttccc	gccgggtgt	ggccctgctt	cgcccagagc	ccctgcaccg	600
ggggacagca	gacaccttc	tcaaccgggt	taagaagctg	ccttgtcaga	tcaccagcta	660
cctgggtggc	cacaccttag	ggcgccggat	gctgtatcca	ggctctgtgt	acctgctgca	720
gaaggccctc	atgcctgtgc	tgctgcaggg	ccaggcccg	ctggtggaag	agtgtaatgg	780
gcgcggggca	aagctgctgg	cctgtgatgg	caatgagatt	gacaccatgt	ttgtggaccg	840
gcgggggaca	gctgagcccc	aggacagaag	ctggtgatct	gctgtgargg	gaatgctggg	900
ttttatgarg	tgggctgcgt	ctccamgcc	ctggaaactg	gatatcatc	ckgggtggaa	960
tcaccagct	ttgctggaac	acagggggt	ccattcccgc	aaaatgaagg	ctaagccat	1020
gg						1022

<210> 120

<211> 2311

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (654)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (2293)

<223> n equals a,t,g, or c

<400> 120

tcgacccacg	cgtecgcttc	ctcatgctgc	ccgctcgctt	accgttcagg	ctgctgagcc	60
tttcccttcg	tggatccgct	cccacggcag	cgcgccatgg	ctccggggag	ccgctcctgg	120
agaggagggtg	cgctgccgcc	tccctcttcc	agcactcatc	gactctggga	cgcgagcttc	180
cttatgaccc	cgtggacacg	gagggctttg	gagaaggttg	tgacatgcag	gagcgttttc	240
tgttcccggg	gtacatcctg	gatccggagc	cgcaaccac	ccgcgaaaaa	cagctgcagg	300
agctccagca	acagcaggag	gaggaggagc	gacagaggca	gcagcgccgg	gaggagcggc	360
gacagcaaaa	cctacggggc	aggtcccggg	agcaccgggt	cgtggggcac	ccggacccgg	420
cattgccggc	cagcggcgctg	aactgctcgg	gctgcggggc	asascgcac	tgccaggacg	480
ccggagtgc	cggtacctg	ccccgagaga	agttccctcc	cacggcggag	gcagacggcg	540
ggctggcacg	gaccgtgtgc	cagcgctgct	ggctgctgtc	gcaccaccgg	cgcgctctac	600
gcctgcagg	gagccgcgag	cagtacctgg	agctgggtgag	cgccgcgttg	cgngggcccg	660
gcccctccct	ggtgctctac	atgggtggacc	tgctggacct	gcccgaacgc	ctgctgcccc	720
acttgccgcg	gctggtgggc	cccaagcagc	tgatcgtgct	gggaaacaaa	gtggaccttc	780
tgccccagga	tgctcctggc	taccggcaga	ggctgcggga	gcgactgtgg	gaggactgtg	840
cccgcgcgg	gctcctgctg	gcccctggca	ccaagggcca	cagcgccccg	tcaaggacga	900
gccacaggac	ggggagaatc	cgaatccgcc	gaactggctc	cgcacagtgg	tcagggacgt	960
gcggctgac	agcgccaaga	ccggctatgg	agtggaaag	ttgatctctg	cccttcagcg	1020
ctcctggcgc	taccgtgggg	acgtctactt	agtgggcgcc	accaacgccg	gcaaatccac	1080
tctctttaac	acgctcctgg	agtcggatta	ctgactgcac	aagggtcccg	aggccatcga	1140
cagagccacc	atctccccct	ggccaggtag	tacattaaac	cttctgaagt	ttcctatttg	1200
caacccaact	ccttacagaa	tgtttaaaag	gcattcaaaga	cttaaaaaag	attcaactca	1260
agctgaagaa	gatcttagtg	agcaagaaca	aaatcagctt	aatgtcctca	aaaagcatgg	1320
ttatgtcgta	ggaagagttg	gaaggacatt	cttgattcca	gaagaacaga	aggataacat	1380
tccctttgag	tttgatgctg	attcacttgc	ctttgacatg	gaaaatgacc	ctgttatggg	1440
tacacacaaa	tccacaaaac	aagtagaatt	gactgcacaa	gatgtgaaag	atgcccactg	1500
gttttatgac	acccctggaa	ttacaaaaa	aaattgtatt	ttaaatcttc	taacagaaaa	1560

agaagtaaat	attgttttgc	caacacagtc	cattgtttcca	agaacttttg	tgcttaaacc	1620
aggaatgggt	ctgtttttgg	gtgctatagg	ccgcatagat	ttcctgcagg	gaaatcagtc	1680
agcttggttt	acagtcgtgg	cttccaacat	cctccctgtg	catatcacct	ccttggtgacag	1740
ggcagacgct	ctgtatcaga	agcatgcagg	tcatacggtta	ctccagattc	caatgggtgg	1800
aaaagaacga	atggcrggat	ttcctcctct	tggtgctgaa	gacattatgt	taaaagaagg	1860
actgggggca	tctgaagcag	tggccgacat	caagttttcc	tctgcagggt	gggtttcagt	1920
aacacctaat	tttaaggaca	gactgcatct	ccgaggctat	acacctgaag	gaacagtttt	1980
gaccgtccgg	ccccctctct	tgccatata	tgtaacatc	aaaggacagc	gcatcaagaa	2040
aagtgtggcc	tataaaacca	agaagcctcc	ttcccttatg	tacaacgtga	ggaagaagaa	2100
aggaaagata	aatgtatgag	accgaccttg	ttcactccag	atattaactg	tattgaacac	2160
aacaaaatac	attgaatttg	tattaaacat	ataacgcata	aataaagctc	ccattcttac	2220
ccttaaaaaa	aaaaaaaaag	ggcggccgct	ctagaggatc	caagcttacg	tacgcgtgca	2280
tgcgacgtca	tancctcgtc	ataggaactg	g			2311

<210> 121

<211> 1286

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (1284)

<223> n equals a,t,g, or c

<400> 121

gggcgcgcgg	gtgaaaggcg	cattgatgca	gcctgcggcg	gcctcggagc	gcggcgggagc	60
agacgctgac	caggttcttc	tcctcgggtc	cctccgcctc	cagctccgcg	ctgcccgcca	120
gccgggagcc	atgcgacccc	agggccccgc	cgctccccg	cagcggctcc	gcggcctcct	180
gctgctcctg	ctgctgcagc	tgcccgcgcc	gtcgagcgcc	tctgagatcc	ccaaggggaa	240
gcaaaaaggc	cactccggca	gagggagggt	gtggacctgt	ataatggaat	gtgcttacaa	300
gggccagcag	gagtgcctgg	tcgagacggg	agccctgggg	ccaatggcat	tcgggttaca	360
cctgggatcc	caggtcggga	tggattcaaa	ggagaaaagg	gggaatgtct	gagggaaagc	420
tttgaggagt	cctggacacc	caactacaag	cagtgttcat	ggagttcatt	gaattatggc	480
atagatcttg	ggaaaattgc	ggagtgtaca	tttacaaga	tgctttcaaa	tagtgctcta	540
agagttttgt	tcagtggctc	acttcggcta	aaatgcagaa	atgcattgctg	tcagcgttgg	600
tatttcacat	tcaatggagc	tgaatgttca	ggacctcttc	ccattgaagc	tataatttat	660
ttggaccaag	gaagccctga	aatgaattca	acaatttaata	ttcatcgcac	ttcttctgtg	720
gaaggacttt	gtgaaggaa	tgggtgctga	ttagtgatg	ttgctatctg	ggttggcact	780
tgttcagatt	acccaaaagg	agatgcttct	actggatgga	attcagtttc	tcgcattcatt	840
attgaagaac	taccaaaata	aatgctttaa	ttttcatttg	ctacctcttc	ttttattatg	900
ccttggaatg	gttcacttaa	atgacatttt	aaataagttt	atgtatacat	ctgaatgaaa	960
agcaaaagta	aatatgttta	cagaccaaag	tgtgatttca	cactgttttt	aaatctagca	1020
ttattcattt	tgcttcaatc	aaaagtgggt	tcaatatttt	ttttagtggg	ttagaatact	1080
ttcttcatag	tcacattctc	tcaacctata	atttgggaata	ttgttggtgt	cttttgtttt	1140
ttctcttagt	atagcatttt	taaaaaaata	taaaagctac	caatctttgt	acaatttgta	1200
aatgctaaga	atttttttta	tatctgttaa	ataaaaatta	ttttcaacaa	aaaaaaaaaa	1260
aaaaaaaaaa	aaaaaaaaaa	aaanaa				1286

<210> 122

<211> 1380

<212> DNA

<213> Homo sapiens

<400> 122

cagaccgcgc	gggcaaacgg	actggggcca	agagccggga	gcgcgggcgc	aaaggcacca	60
gggcccgcgc	agggcgccgc	gcacacggcc	ttgggggttc	tgccggcctt	cggggtgcgc	120
tctgcctct	agccatgggg	tccgcagcgt	tggagatcct	ggcctgggtg	ctgtgcctgg	180

tgggctgggg	gggtctgac	ctggcgtg	ggctgccc	gtggcaggtg	accgccttcc	240
tggaccacaa	catcgtgacg	gcgcagacca	cctggaagg	gctgtggatg	tcgtgctg	300
tgcagagcac	gggcacatgc	agtgc aaagt	gtacgactcg	gtgctggctc	tgagcaccga	360
ggtgcaggcg	gcgcggggcg	tcaccgtgag	cgccgtgctg	ctggcgttcg	ttgcgctctt	420
cgtgaccctg	gcggggcg	agtgcaccac	ctgcctggcc	ccggggcccg	ccaaggcg	480
tgtggccctc	acgggaggcg	tgctctacct	gttttgcggg	ctgctggcgc	tcgtgccact	540
ctgctggttc	gccaaacattg	tcgtccgcga	gttttacgac	ccgtctgtgc	ccgtgtcgca	600
gaagtacgag	ctggggcg	gctgtacatc	ggctgggcgg	ccaccgcgct	gctcatggta	660
ggcggctgcc	tcttgtgctg	cggcgcctgg	gtctgcaccg	gccgtcccga	cctcagcttc	720
cccgtgaagt	actcagcgcc	gcggcgggccc	acggccaccg	gcgactacga	caagaagaac	780
tacgtctgag	ggcgctgggc	acggccggggc	ccctcctgcc	agccacgcct	gcgaggcggt	840
ggataagcct	ggggagcccc	gc atggaccg	cggtctccgc	cggttagcgc	ggcgcgagg	900
ctcctcgga	cgctcggctc	tgccgccga	cgcggtcct	ggatccgctc	ctgcctgcgc	960
ccgcagctga	ccttctctcg	ccactagccc	ggccctgccc	ttaacagacg	gaatgaagtt	1020
tcctttctg	tgccggcg	tgtttccata	ggcagagcgg	gtgtcagact	gaggatttcg	1080
cttccctcc	aagacgctgg	gggtcttggc	tgctgcctta	cttcccagag	gctcctgctg	1140
acttcggagg	ggcggtgca	gagcccagg	ccccccaccg	aagatgtgta	cacctggtct	1200
ttactccatc	ggcaggggccc	gagcccagg	accagtgact	tggcctggac	ctcccgggtct	1260
cactccagca	tctccccagg	caaggcttgt	gggcaccgga	gcttgagaga	ggcgggaggt	1320
gggaaggcta	agaatctgct	tagtaaatgg	tttgaactct	caaaaaaaaa	aaaaaaaaaa	1380

<210> 123

<211> 3793

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (1102)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1132)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1199)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1228)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1229)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (1231)

<223> n equals a,t,g, or c

<220>

<221> SITE